Journal of Medical Internet Research

Impact Factor (2018): 4.945 - ranked #1 medical informatics journal by Impact Factor
Volume 23 (2021), Issue 8 ISSN: 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Reviews

eHealth in Geriatric Rehabilitation: Systematic Review of Effectiveness, Feasibility, and Usability (e24015)
Jules Kraaijkamp, Eléonore van Dam van Isselt, Anke Persoon, Anke Versluis, Niels Chavannes, Wilco Achterberg. 9

Overview and Strategy Analysis of Technology-Based Nonpharmacological Interventions for In-Hospital Delirium Prevention and Reduction: Systematic Scoping Review (e26079)
Chan Kim, Esther van der Heide, Thomas van Rompay, Gijsbertus Verkerke, Geke Ludden. 29

The Influence of Web-Based Tools on Maternal and Neonatal Outcomes in Pregnant Adolescents or Adolescent Mothers: Mixed Methods Systematic Review (e26786)
Jania Wu, Nurulhuda Ahmad, Miny Samuel, Susan Logan, Citra Mattar. 51

Hospital Investment Decisions in Healthcare 4.0 Technologies: Scoping Review and Framework for Exploring Challenges, Trends, and Research Directions (e27571)
Roberto Vassolo, Alejandro Mac Cawley, Guilherme Tortorella, Flavio Fogliatto, Diego Tlapa, Gopalakrishnan Narayanamurthy. 65

Diagnostic Accuracy of Smartwatches for the Detection of Cardiac Arrhythmia: Systematic Review and Meta-analysis (e28974)
Scarlet Nazarian, Kyle Lam, Ara Darzi, Hutan Ashrafian. 87

Ecological Momentary Assessment and mHealth Interventions Among Men Who Have Sex With Men: Scoping Review (e27751)
Viktor Clark, Sunny Kim. 220

Telehealth Interventions and Outcomes Across Rural Communities in the United States: Narrative Review (e29575)
Michael Butzner, Yendelela Cuffee. 449

Applications of Extended Reality in Ophthalmology: Systematic Review (e24152)
Chee Ong, Marcus Tan, Michael Lam, Victor Koh. 708

Measurement Properties of Patient-Reported Outcome Measures for Diabetes: Systematic Review (e25002)
Priscilla Wee, Yu Kwan, Dionne Loh, Jie Phang, Troy Puar, Truls Østbye, Julian Thumboo, Sungwon Yoon, Lian Low. 764

Computer-Aided Diagnosis of Diminutive Colorectal Polyps in Endoscopic Images: Systematic Review and Meta-analysis of Diagnostic Test Accuracy (e29682)
Chang Bang, Jae Lee, Gwang Baik. 846

Blockchain Technology Projects to Provide Telemedical Services: Systematic Review (e17475)
Konstantin Koshechkin, Georgy Lebedev, George Radzievsky, Ralf Seepold, Natividad Martinez. 950

Viewpoints

Every Child on the Map: A Theory of Change Framework for Improving Childhood Immunization Coverage and Equity Using Geospatial Data and Technologies (e29759)
Sarah Chaney, Patricia Mechael, Nay Thu, Mamadou Diallo, Carine Gachen. 102

Podcasts for the Delivery of Medical Education and Remote Learning (e29168)
Julliana Newman, Andrew Liew, Jon Bowles, Kelly Soady, Steven Inglis. 111

Original Papers

Web-Based Guidance Through Assisted Reproductive Technology (myFertiCare): Patient-Centered App Development and Qualitative Evaluation (e25389)
Ellen Sparidaens, Rosella Hermens, Didi Braat, Willianne Nelen, Kathrin Fleischer. 118

Evaluation of a Digital Health Initiative in Illicit Substance Use: Cross-sectional Survey Study (e29026)
Steph Kershaw, Louise Birrell, Hannah Deen, Nicola Newton, Lexine Stapinski, Katrina Champion, Frances Kay-Lambkin, Maree Teesson, Cath Chapman. 128

Effectiveness of Internet-Based Cognitive Behavior Therapy (Fatigue in Teenagers on the Internet) for Adolescents With Chronic Fatigue Syndrome in Routine Clinical Care: Observational Study (e24839)
Eline Albers, Linde Nijhof, Emma Berkelbach van der Sprenkel, Elise van de Putte, Sanne Nijhof, Hans Knoop. 143

New Internet-Based Warfarin Anticoagulation Management Approach After Mechanical Heart Valve Replacement: Prospective, Multicenter, Randomized Controlled Trial (e29529)
Zhihui Zhu, Chenyu Li, Jinglun Shen, Kaisheng Wu, Yuehuan Li, Kun Liu, Fan Zhang, Zhenhua Zhang, Yan Li, Jie Han, Ying Qin, Yu Yang, Guangpu Fan, Huajun Zhang, Zheng Ding, Dong Xu, Yu Chen, Yingli Zheng, Zhe Zheng, Xu Meng, Haibo Zhang. 153

Internet-Based HIV Self-Testing Among Men Who Have Sex With Men Through Pre-exposure Prophylaxis: 3-Month Prospective Cohort Analysis From China (e23978)
Jing Zhang, Joseph Tucker, Weiming Tang, Hongyi Wang, Zhenxing Chu, Qinghai Hu, Xiaojie Huang, Yaokai Chen, Hui Wang, Xiaoqing He, Yao Li, Lukun Zhang, Zhili Hu, Rantong Bao, Shangcao Li, Hang Li, Haibo Ding, Yongjun Jiang, Wenqing Geng, Junjie Xu, Hong Shang. 165

The Effectiveness of a Web-Based Self-Help Program to Reduce Alcohol Use Among Adults With Drinking Patterns Considered Harmful, Hazardous, or Suggestive of Dependence in Four Low- and Middle-Income Countries: Randomized Controlled Trial (e21686)
Michael Schaub, Marcela Tiburcio, Nora Martínez-Vélez, Atul Ambekar, Roshan Bhad, Andreas Wenger, Christian Baumgartner, Dzianis Padruchny, Sergey Osipchik, Vladimir Poznyak, Dag Rekve, Fabricio Landi Moraes, André Monezi Andrade, Maria Oliveira Souza-Formigoni, WHO E-Health Project On Alcohol And Health Investigators Group. 178

Informing Patients With Esophagogastric Cancer About Treatment Outcomes by Using a Web-Based Tool and Training: Development and Evaluation Study (e27824)
Loïs van de Water, Héctor van den Boorn, Florian Hoxha, Inge Henselmans, Mart Calff, Mirjam Sprangers, Ameen Abu-Hanna, Ellen Smets, Hanneke van Laarhoven. 190

Effects of the Interactive Web-Based Video “Mon Coeur, Mon BASIC” on Drug Adherence of Patients With Myocardial Infarction: Randomized Controlled Trial (e21938)
Christel Bruggmann, Julien Adjedj, Sylvain Sardy, Olivier Muller, Pierre Voirol, Farshid Sadeghipour. 205

Gaps and Future Challenges of Italian Apps for Pregnancy and Postnatal Care: Systematic Search on App Stores (e29151)

Laura Brunelli, Chiara De Vita, Fabrizio Cenedese, Michela Cinello, Marta Paris, Francesca Samogizio, Anja Starec, Michele Bava, Margherita Dal Cin, Sara Zanchiello, Tamara Stampalija. 229

Feasibility of an Interactive Health Coaching Mobile App to Prevent Malnutrition and Muscle Loss in Esophageal Cancer Patients Receiving Neoadjuvant Concurrent Chemoradiotherapy: Prospective Pilot Study (e28695)

Kyungmi Yang, Dongryul Oh, Jae Noh, Han Yoon, Jong-Mu Sun, Hong Kim, Jae Zo, Young Shim, Hyunyoung Ko, Jungeun Lee, Youngin Kim.
2 3 9

Perceptions of Alerts Issued by Social Media Platforms in Response to Self-injury Posts Among Latinx Adolescents: Qualitative Analysis (e28931)

Linnea Laestadius, Katherine Craig, Celeste Campos-Castillo. 251

Psychological Violence Against Arab Women in the Context of Social Media: Web-Based Questionnaire Study (e27944)

Ahmed Omar Bali, Emad Omer, Kawa Abdulridha, Araz Ahmad. 260

Association Between Social Media Use and Cancer Screening Awareness and Behavior for People Without a Cancer Diagnosis: Matched Cohort Study (e26395)

Lei Qin, Xiaomei Zhang, Anlin Wu, James Miser, Yen-Lin Liu, Jason Hsu, Ben-Chang Shia, Linglong Ye. 270

Distant Supervision for Mental Health Management in Social Media: Suicide Risk Classification System Development Study (e26119)

Guanghui Fu, Changwei Song, Jianqiang Li, Yue Ma, Pan Chen, Ruiqian Wang, Bing Yang, Zhisheng Huang. 280

Factors That Help and Hinder the Implementation of Digital Depression Prevention Programs: School-Based Cross-sectional Study (e26223)

Joanne Beames, Lara Johnston, Bridianne O'Dea, Michelle Torok, Helen Christensen, Katherine Boydell, Aliza Werner-Seidler. 293

Using Machine Learning–Based Approaches for the Detection and Classification of Human Papillomavirus Vaccine Misinformation: Infodemiology Study of Reddit Discussions (e26478)

Jingcheng Du, Sharice Preston, Hanxiao Sun, Ross Shegog, Rachel Cunningham, Julie Boom, Lara Savas, Muhammad Amith, Cui Tao. . .
3 0 7

Mining of Opinions on COVID-19 Large-Scale Social Restrictions in Indonesia: Public Sentiment and Emotion Analysis on Online Media (e28249)

Andi Tri Sakti, Emma Mohamad, Arina Azlan. 319

Forecasting the COVID-19 Epidemic by Integrating Symptom Search Behavior Into Predictive Models: Infoveillance Study (e28876)

Alessandro Rabiolo, Eugenio Alladio, Esteban Morales, Andrew McNaught, Francesco Bandello, Abdelmonem Afifi, Alessandro Marchese. . .
3 3 2

Using Infodemiology Metrics to Assess Public Interest in Liver Transplantation: Google Trends Analysis (e21656)

Maria Effenberger, Andreas Kronbichler, Erica Bettac, Felix Grabherr, Christoph Grander, Timon Adolph, Gert Mayer, Heinz Zoller, Paul Perco, Herbert Tilg. 343

Development of a COVID-19 Web Information Transmission Structure Based on a Quadruple Helix Model: Webometric Network Approach Using Bing (e27681)

Yu Zhu, Han Park. 356

Online Public Attention Toward Premature Ejaculation in Mainland China: Infodemiology Study Using the Baidu Index (e30271)

Shanzun Wei, Ming Ma, Xi Wen, Changjing Wu, Guonian Zhu, Xiangfu Zhou. 375

Exploring the Expression Differences Between Professionals and Laypeople Toward the COVID-19 Vaccine: Text Mining Approach (e30715)

Chen Luo, Kaiyuan Ji, Yulong Tang, Zhiyuan Du. 386

Leveraging Multimedia Patient Engagement to Address Minority Cerebrovascular Health Needs: Prospective Observational Study (e28748)

Elizabeth Noser, Jing Zhang, Mohammad Rahbar, Anjail Sharrief, Andrew Barreto, Sandi Shaw, James Grotta, Sean Savitz, Nneka Ifejika. . .
4 0 2

Patients With Cancer Searching for Cancer- or Health-Specific Web-Based Information: Performance Test Analysis (e23367)

Lukas Lange-Drenth, Holger Schulz, Gero Endsin, Christiane Bleich. 414

Investigating the Use of Telemedicine for Digitally Mediated Delegation in Team-Based Primary Care: Mixed Methods Study (e28151)

Michael Knop, Marius Mueller, Bjoern Niehaves. 428

Using a Smartphone Application for the Accurate and Rapid Diagnosis of Acute Anterior Intracranial Arterial Occlusion: Usability Study (e28192)

Teppei Komatsu, Kenichiro Sakai, Yasuyuki Iguchi, Hiroyuki Takao, Toshihiro Ishibashi, Yuichi Murayama. 458

A Home-Based eHealth Intervention for an Older Adult Population With Food Insecurity: Feasibility and Acceptability Study (e26871)

Luís Gomes, Maria Gregório, Tatiana Iakovleva, Rute Sousa, John Bessant, Pedro Oliveira, Jaime Branco, Helena Canhão, Ana Rodrigues. .
4 6 7

Patient and Professional Experiences With Virtual Antenatal Clinics During the COVID-19 Pandemic in a UK Tertiary Obstetric Hospital: Questionnaire Study (e25549)

Lauren Quinn, Oluwafumbi Olajide, Marsha Green, Hazem Sayed, Humera Ansar. 483

Comparison of the Differences Between Web-Based and Traditional Questionnaire Surveys in Pediatrics: Comparative Survey Study (e30861)

Heping Fang, Ruoling Xian, Zhuoying Ma, Mingyue Lu, Yan Hu. 492

Health Information Technology Use Among Persons With Self-reported Atherosclerotic Cardiovascular Disease: Analysis of the 2011-2018 National Health Interview Survey (e23765)

Uchenna Nwokeji, Erin Spaulding, Rongzi Shan, Ruth-Alma Turkson-Ocran, Diana Baptiste, Binu Koirala, Timothy Plante, Seth Martin, Yvonne Commodore-Mensah. 498

Obesity and BMI Cut Points for Associated Comorbidities: Electronic Health Record Study (e24017)

Natalie Liu, Jen Birstler, Manasa Venkatesh, Lawrence Hanrahan, Guanhua Chen, Luke Funk. 511

A Fine-Tuned Bidirectional Encoder Representations From Transformers Model for Food Named-Entity Recognition: Algorithm Development and Validation (e28229)

Riste Stojanov, Gorjan Popovski, Gjorgjina Cenikj, Barbara Korouši Seljak, Tome Eftimov. 524

Automated Screening for Social Anxiety, Generalized Anxiety, and Depression From Objective Smartphone-Collected Data: Cross-sectional Study (e28918)

Daniel Di Matteo, Kathryn Fotinos, Sachinthya Lokuge, Geneva Mason, Tia Sternat, Martin Katzman, Jonathan Rose. 537

Next-Generation Digital Biomarkers for Tuberculosis and Antibiotic Stewardship: Perspective on Novel Molecular Digital Biomarkers in Sweat, Saliva, and Exhaled Breath (e25907)

Noe Brasier, Michael Osthoff, Fiorangelo De Ieso, Jens Eckstein. 551

Classification of Children With Autism and Typical Development Using Eye-Tracking Data From Face-to-Face Conversations: Machine Learning Model Development and Performance Evaluation (e29328)
Zhong Zhao, Haiming Tang, Xiaobin Zhang, Xingda Qu, Xinyao Hu, Jianping Lu. 561

Anticipated Benefits and Concerns of Sharing Hospital Outpatient Visit Notes With Patients (Open Notes) in Dutch Hospitals: Mixed Methods Study (e27764)
Sharon Janssen, Nynke Venema-Taat, Stephanie Medlock. 575

Effects of Patient Portal Use on Patient Satisfaction: Survey and Partial Least Squares Analysis (e19820)
Aaron Kinney, Balaji Sankaranarayanan. 589

Mental Health Care Professionals’ Appraisal of Patients’ Use of Web-Based Access to Their Electronic Health Record: Qualitative Study (e28045)
Antonius van Rijt, Pauline Hulter, Anne Weggelaar-Jansen, Kees Ahaus, Bettine Pluut. 605

The Application of a Case-Based Social Media–Assisted Teaching Method in Cariology Education Comparative Study (e29372)
Li Li, Xiaobin Liu, Zeyuan Chen, Liyuan Wang, Xiaoli Lian, Huiru Zou. 617

The Impact of Gamification-Induced Users' Feelings on the Continued Use of mHealth Apps: A Structural Equation Model With the Self-Determination Theory Approach (e24546)
Tong Wang, Lingye Fan, Xu Zheng, Wei Wang, Jun Liang, Kai An, Mei Ju, Jianbo Lei. 628

The Use and Effectiveness of an Online Diagnostic Support System for Blood Film Interpretation: Comparative Observational Study (e20815)
Claire Hutchinson, Michelle Brereton, Julie Adams, Barbara De La Salle, Jon Sims, Keith Hyde, Richard Chasty, Rachel Brown, Karen Rees-Unwin, John Burthem. 643

Development and Validation of Unplanned Extubation Prediction Models Using Intensive Care Unit Data: Retrospective, Comparative, Machine Learning Study (e23508)
Sujeong Hur, Ji Min, Junsang Yoo, Kyunga Kim, Chi Chung, Patricia Dykes, Won Cha. 656

A Modified Public Health Automated Case Event Reporting Platform for Enhancing Electronic Laboratory Reports With Clinical Data: Design and Implementation Study (e26388)
Ninad Mishra, Jon Duke, Saugat Karki, Myung Choi, Michael Riley, Andrey Ilatovskiy, Marla Gorges, Leslie Lenert. 670

Interorganizational Knowledge Sharing to Establish Digital Health Learning Ecosystems: Qualitative Evaluation of a National Digital Health Transformation Program in England (e23372)
Kathrin Cresswell, Aziz Sheikh, Bryony Franklin, Marta Krasuska, Hung The Nguyen, Susan Hinder, Wendy Lane, Hajar Mozaffar, Kathy Mason, Sally Eason, Henry Potts, Robin Williams. 680

Patterns of eHealth Website User Engagement Based on Cross-site Clickstream Data: Correlational Study (e29299)
Jia Li, Kanghui Yu, Xinyu Bao, Xuan Liu, Junping Yao. 692

Variation of Daily Care Demand in Swiss General Hospitals: Longitudinal Study on Capacity Utilization, Patient Turnover and Clinical Complexity Levels (e27163)
Narayan Sharma, René Schwendimann, Olga Endrich, Dietmar Ausserhofer, Michael Simon. 725

Alignment of Key Stakeholders’ Priorities for Patient-Facing Tools in Digital Health: Mixed Methods Study (e24890)
Courtney Lyles, Julia Adler-Milstein, Crishyashi Thao, Sarah Lisker, Sarah Nouri, Urmimala Sarkar. 740

The Effect of Collaborative Reviews of Electronic Patient-Reported Outcomes on the Congruence of Patient- and Clinician-Reported Toxicity in Cancer Patients Receiving Systemic Therapy: Prospective, Multicenter, Observational Clinical Trial (e29271)
Andreas Trojan, Nicolas Leuthold, Christoph Thomssen, Achim Rody, Thomas Winder, Andreas Jakob, Claudine Egger, Ulrike Held, Christian Jackisch. 752

Patient and Clinician Characteristics Associated With Secure Message Content: Retrospective Cohort Study (e26650)
Dawn Heisey-Grove, Cheryl Rathert, Laura McClelland, Kevin Jackson, Jonathan DeShazo. 790

Artificial Intelligence–Based Prediction of Lung Cancer Risk Using Nonimaging Electronic Medical Records: Deep Learning Approach (e26256)
Marvin Yeh, Yu-Hsiang Wang, Hsuan-Chia Yang, Kuan-Jen Bai, Hsiao-Han Wang, Yu-Chuan Li. 804

Construction of Genealogical Knowledge Graphs From Obituaries: Multitask Neural Network Extraction System (e25670)
Kai He, Lixia Yao, JiaWei Zhang, Yufei Li, Chen Li. 818

An Artificial Neural Network Prediction Model for Posttraumatic Epilepsy: Retrospective Cohort Study (e25090)
Xueping Wang, Jie Zhong, Ting Lei, Deng Chen, Haijiao Wang, Lina Zhu, Shanshan Chu, Ling Liu. 832

Patient Perceptions on Data Sharing and Applying Artificial Intelligence to Health Care Data: Cross-sectional Survey (e26162)
Ravi Aggarwal, Soma Farag, Guy Martin, Hutan Ashrafian, Ara Darzi. 867

A Machine Learning Approach to Passively Informed Prediction of Mental Health Risk in People with Diabetes: Retrospective Case-Control Analysis (e27709)
Jessica Yu, Carter Chiu, Yajuan Wang, Eldin Dzubur, Wei Lu, Julia Hoffman. 879

Image Processing for Public Health Surveillance of Tobacco Point-of-Sale Advertising: Machine Learning–Based Methodology (e24408)
Ned English, Andrew Anesetti-Rothermel, Chang Zhao, Andrew Latterner, Adam Benson, Peter Herman, Sherry Emery, Jordan Schneider, Shyanika Rose, Minal Patel, Barbara Schillo. 891

Predicting Kidney Graft Survival Using Machine Learning Methods: Prediction Model Development and Feature Significance Analysis Study (e26843)
Syed Naqvi, Karthik Tennankore, Amanda Vinson, Patrice Roy, Syed Abidi. 904

Real-Time Respiratory Tumor Motion Prediction Based on a Temporal Convolutional Neural Network: Prediction Model Development Study (e27235)
Panchun Chang, Jun Dang, Jianrong Dai, Wenzheng Sun. 925

Perspectives on Fruit and Vegetable Consumption and Government Dietary Guidelines: Content Analysis of Comments on News Websites (e19917)
B Hornsby, H Ensaff. 937

Chinese Patients’ Intention to Use Different Types of Internet Hospitals: Cross-sectional Study on Virtual Visits (e25978)
Liyun Liu, Lizheng Shi. 961

Integrating Welfare Technology in Long-term Care Services: Nationwide Cross-sectional Survey Study (e22316)
Hanne Rostad, Randi Stokke. 975

Digital Orientation of Health Systems in the Post–COVID-19 “New Normal” in the United States: Cross-sectional Survey (e30453)

Jiban Khuntia, Xue Ning, Rulon Stacey. 991

Characterization of Global Research Trends and Prospects on Single-Cell Sequencing Technology: Bibliometric Analysis (e25789)

Quan Wang, Ke-Lu Yang, Zhen Zhang, Zhu Wang, Chen Li, Lun Li, Jin-Hui Tian, Ying-Jiang Ye, Shan Wang, Ke-Wei Jiang. 1004

Evaluating Epidemiological Risk by Using Open Contact Tracing Data: Correlational Study (e28947)

Stefano Piotto, Luigi Di Biasi, Francesco Marrafino, Simona Concilio. 1020

Association Between Public Opinion and Malaysian Government Communication Strategies About the COVID-19 Crisis: Content Analysis of Image Repair Strategies in Social Media (e28074)

Nasaai Masngut, Emma Mohamad. 1031

Exploring Changes to the Actionability of COVID-19 Dashboards Over the Course of 2020 in the Canadian Context: Descriptive Assessment and Expert Appraisal Study (e30200)

Erica Barbazza, Damir Ivankovi , Sophie Wang, Kendall Gilmore, Mircha Poldrugovac, Claire Willmington, Nicolas Larrain, Véronique Bos, Sara Allin, Niek Klazinga, Dionne Kringos. 1040

Self-Focused and Other-Focused Health Concerns as Predictors of the Uptake of Corona Contact Tracing Apps: Empirical Study (e29268)

Fenne große Deters, Tabea Meier, Anne Milek, Andrea Horn. 1054

Leveraging Transfer Learning to Analyze Opinions, Attitudes, and Behavioral Intentions Toward COVID-19 Vaccines: Social Media Content and Temporal Analysis (e30251)

Siru Liu, Jili Li, Jialin Liu. 1069

Long-term Effects of the COVID-19 Pandemic on Public Sentiments in Mainland China: Sentiment Analysis of Social Media Posts (e29150)

Hao Tan, Sheng-Lan Peng, Chun-Peng Zhu, Zuo You, Ming-Cheng Miao, Shu-Guang Kuai. 1083

Predictors of COVID-19 Preventive Perceptions and Behaviors Among Millennials: Two Cross-sectional Survey Studies (e30612)

Christopher Beaudoin, Traci Hong. 1095

Digital Surveillance Through an Online Decision Support Tool for COVID-19 Over One Year of the Pandemic in Italy: Observational Study (e29556)

Alberto Tozzi, Francesco Gesualdo, Emanuele Urbani, Alessandro Sbenaglia, Roberto Ascione, Nicola Procopio, Ileana Croci, Caterina Rizzo.
1 1 1 0

Changes in Public Response Associated With Various COVID-19 Restrictions in Ontario, Canada: Observational Infoveillance Study Using Social Media Time Series Data (e28716)

Antony Chum, Andrew Nielsen, Zachary Bellows, Eddie Farrell, Pierre-Nicolas Durette, Juan Banda, Gerald Cupchik. 1120

Implementation of Digital Monitoring Services During the COVID-19 Pandemic for Patients With Chronic Diseases: Design Science Approach (e24181)

Luís Lapão, Mariana Peyroteo, Melanie Maia, Jorge Seixas, João Gregório, Miguel Mira da Silva, Bruno Heleno, Jorge Correia. 1137

Use of Iris Scanning for Biometric Recognition of Healthy Adults Participating in an Ebola Vaccine Trial in the Democratic Republic of the Congo: Mixed Methods Study (e28573)

Trésor Zola Matuvanga, Ginger Johnson, Ynke Larivière, Emmanuel Esanga Longomo, Junior Matangila, Vivi Maketa, Bruno Lapika, Patrick Mitashi, Paula Mc Kenna, Jessie De Bie, Jean-Pierre Van Geertruyden, Pierre Van Damme, Hypolite Muhindo Mavoko. 1147

Lateralization and Bodily Patterns of Segmental Signs and Spontaneous Pain in Acute Visceral Disease: Observational Study (e27247)
Nour Shaballout, Anas Aloumar, Jorge Manuel, Marcus May, Florian Beissner. 1156

Corrigenda and Addenda

Correction: The State of Evidence in Patient Portals: Umbrella Review (e32421)
Marcy Antonio, Olga Petrovskaya, Francis Lau. 1018

Review

eHealth in Geriatric Rehabilitation: Systematic Review of Effectiveness, Feasibility, and Usability

Jules J M Kraaijkamp[1,2], MSc; Eléonore F van Dam van Isselt[1], MD, PhD; Anke Persoon[3], PhD; Anke Versluis[1], PhD; Niels H Chavannes[1], MD, PhD; Wilco P Achterberg[1], MD, PhD

[1]Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, Netherlands
[2]ZZG Zorggroep, Nijmegen, Netherlands
[3]Department of Primary and Community Care, Radboud University Medical Center Nijmegen, Nijmegen, Netherlands

Corresponding Author:
Jules J M Kraaijkamp, MSc
Department of Public Health and Primary Care
Leiden University Medical Center
Hippocratespad 21
Leiden, 2333 ZD
Netherlands
Phone: 31 610599869
Email: J.J.M.Kraaijkamp@lumc.nl

Abstract

Background: eHealth has the potential to improve outcomes such as physical activity or balance in older adults receiving geriatric rehabilitation. However, several challenges such as scarce evidence on effectiveness, feasibility, and usability hinder the successful implementation of eHealth in geriatric rehabilitation.

Objective: The aim of this systematic review was to assess evidence on the effectiveness, feasibility, and usability of eHealth interventions in older adults in geriatric rehabilitation.

Methods: We searched 7 databases for randomized controlled trials, nonrandomized studies, quantitative descriptive studies, qualitative research, and mixed methods studies that applied eHealth interventions during geriatric rehabilitation. Included studies investigated a combination of effectiveness, usability, and feasibility of eHealth in older patients who received geriatric rehabilitation, with a mean age of ≥70 years. Quality was assessed using the Mixed Methods Appraisal Tool and a narrative synthesis was conducted using a harvest plot.

Results: In total, 40 studies were selected, with clinical heterogeneity across studies. Of 40 studies, 15 studies (38%) found eHealth was at least as effective as non-eHealth interventions (56% of the 27 studies with a control group), 11 studies (41%) found eHealth interventions were more effective than non-eHealth interventions, and 1 study (4%) reported beneficial outcomes in favor of the non-eHealth interventions. Of 17 studies, 16 (94%) concluded that eHealth was feasible. However, high exclusion rates were reported in 7 studies of 40 (18%). Of 40 studies, 4 (10%) included outcomes related to usability and indicated that there were certain aging-related barriers to cognitive ability, physical ability, or perception, which led to difficulties in using eHealth.

Conclusions: eHealth can potentially improve rehabilitation outcomes for older patients receiving geriatric rehabilitation. Simple eHealth interventions were more likely to be feasible for older patients receiving geriatric rehabilitation, especially, in combination with another non-eHealth intervention. However, a lack of evidence on usability might hamper the implementation of eHealth. eHealth applications in geriatric rehabilitation show promise, but more research is required, including research with a focus on usability and participation.

(J Med Internet Res 2021;23(8):e24015) doi:10.2196/24015

KEYWORDS

geriatric rehabilitation; eHealth; mHealth; digital health; effectiveness; feasibility; usability; systematic review

Introduction

The world's population is aging rapidly. Currently, 143 million people are aged 80 years or older, and this number is expected to rise to around 426 million in 2050 [1]. Although many older adults are relatively fit, functional decline, multimorbidity, and geriatric syndromes such as frailty or falls are common in older adults [2,3]. A combination of these age-associated conditions triggers an increased risk of adverse outcomes such as hospitalization, functional impairments, and even mortality [4]. Postacute care such as geriatric rehabilitation aims to diminish these age-associated risks. Evidence shows that geriatric rehabilitation can improve functional outcomes and reduce nursing home admissions and mortality [5,6]. On the other hand, the rapidly aging populations and lack of staff are putting pressure on the quality, accessibility, and affordability of geriatric rehabilitation. In regard to these problems, the use of eHealth can be seen as important and promising, as it has the potential to simultaneously improve both rehabilitation outcomes and efficiency.

eHealth can be defined as "the use of digital information and communication to support and/or improve health and health care" [7]. Some examples of eHealth are video communication, exergames (ie, active video games), and mobile apps. Although eHealth can be seen as important and promising, successful implementation of eHealth interventions in geriatric rehabilitation is complex, can be time consuming, and involves a variety of determinants on multiple levels [8-10]. To safely and successfully implement eHealth in geriatric rehabilitation, scientific evaluation of eHealth is key [11,12]. Three important outcome measures for the evaluation of eHealth in geriatric rehabilitation can be identified: effectiveness, feasibility, and usability [9,13].

In terms of effectiveness, previous reviews show that eHealth can improve physical activity, gait, and balance in community-dwelling older adults [14-17]. However, the evidence on effective eHealth in geriatric rehabilitation is scarce and fragmented. To our knowledge, no prior reviews have examined the effectiveness of eHealth in geriatric rehabilitation.

To better understand how eHealth can be used safely, feasibility testing is an important first step [18,19]. The aim of feasibility testing is to "determine whether an intervention is appropriate for further testing" [20,21], but a general accepted standard on feasibility testing is lacking. Examples of factors that can be addressed in feasibility testing are adverse events, adherence, and acceptability [10].

Additionally, usable eHealth is also an important prerequisite for successful implementation [13,19,22]. Usability can be defined as "the extent to which a system, product, or service can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" [23]. For older adults receiving geriatric rehabilitation, usability is especially crucial, since there are certain age-related barriers that may hamper the usability of eHealth [24-26]. These barriers can be categorized into 4 patient-related domains: cognition, physical ability, perception, and motivation [27]. For example, poor vision can make it harder to distinguish certain icons on screens, or cognitive impairment might lead to problems understanding certain eHealth interventions. Often, eHealth is insufficiently tailored to these age-related barriers [28].

Therefore, a systematic review of eHealth in geriatric rehabilitation including the concepts feasibility, usability, and effectiveness was needed. This systematic review can help speed up the implementation process of eHealth and ensure successful adoption of eHealth overall. The aim of this review was to assess evidence on the effectiveness, feasibility, and usability of eHealth interventions in older adults in geriatric rehabilitation.

Methods

Study Registration and Protocol

This systematic review is registered at PROSPERO, with registration number CRD42019133192 [29]. This systematic review was based on the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) statement, which is an evidence-based minimum set of items used for reporting in systematic reviews and meta-analyses [30]. The complete checklist for this review can be found in Multimedia Appendix 1.

Types of Studies and Participants

In this review, we included randomized controlled trials, nonrandomized studies, quantitative descriptive studies, qualitative research, and mixed methods studies. We excluded systematic reviews, abstracts, editorials, and non-English and nonpeer-reviewed studies. Studies were included that examined older patients with a mean age of ≥70 years who received geriatric rehabilitation, which is in line with consensus statements on the organization and delivery of geriatric rehabilitation across Europe [31]. Because there is variability between countries' health care systems and consequently also between countries' provisions of geriatric rehabilitation [31,32], we included studies in different types of settings such as (geriatric) rehabilitation centers, skilled nursing facilities, hospitals, or ambulatory settings. Studies that included patients with a chronic disease with no acute functional decline were excluded.

Interventions and Outcomes

Studies investigated eHealth interventions applied during postacute geriatric rehabilitation. Outcome measures related to the effectiveness of interventions were included if they could be classified based on the World Health Organization's International Classification of Functioning, Disability, and Health (ICF) model [33], which covers the following domains: body functions and structure, activities, participation, environmental factors, and personal factors. For the purpose of this review, we chose to specify feasibility within the following domains: adverse events, adherence, and exclusion rates. Usability outcome measures were classified based on the MOLD-US framework, which is an evidence-based framework of aging barriers that influence the usability of eHealth in older adults and includes 4 categories: cognition, motivation, physical ability, and perception [27]. We included both primary and secondary outcome measures.

Sources and Search Strategy

On March 9, 2019, March 10, 2019, and January 11, 2021, we searched the following databases: MEDLINE, PsycINFO, EMBASE, EMCARE, Cochrane Library, Web of Science, and Central databases. For this review, 3 separate search strings were compiled. The first focused on the effectiveness, the second focused on the feasibility, and the third focused on the usability of eHealth interventions in geriatric rehabilitation. The search string focusing on effectiveness included keywords related to older adults, rehabilitation, and eHealth interventions. Studies were identified when at least 2 of 3 keywords were present. The search strings focusing on feasibility and usability included an additional keyword related to feasibility or usability. In both search strings, keywords were combined using MeSH terms using the Boolean operations "or" and "and." The complete search strings can be found in Multimedia Appendix 2.

Selection of Studies and Data Extraction

We first screened titles of the identified studies. The abstracts of all potentially relevant studies were then screened by 2 authors independently. Next, full texts were obtained and reviewed by the same authors. We excluded studies that did not meet the inclusion criteria. Disagreements between the 2 authors were discussed until a consensus was reached. If a disagreement could not be resolved, a third reviewer was consulted. Data extraction was performed using Covidence, which is an online systematic review management tool [34]. In Covidence, a data extraction form was constructed that included details of publication (ie, author, year, title, country of study, and funding), study design, methods (ie, inclusion and exclusion criteria, population, randomization, statistical analysis, and outcome measures), sample characteristics (ie, age, number of participants, gender, and diagnosis), eHealth intervention (ie, name of intervention, goal of intervention, delivery of intervention, and application of intervention), and primary and secondary outcomes. As the complexity of eHealth interventions influences implementation, we sorted eHealth interventions ranging from simple (ie, video communications, health sensors, or gateways) to complex (ie, robotics, exergames, or virtual reality) [9,35]. One author then extracted the data. A subset of the data (10% of included studies) was also extracted by a second author to check interrater reliability.

Quality Appraisal

The quality of included studies was assessed using the Mixed Methods Appraisal Tool (MMAT) [36], which allowed quality assessment across different study designs. The MMAT is a critical appraisal tool specifically designed to assess the quality of 5 types of study designs: qualitative research, randomized controlled trials, nonrandomized studies, quantitative descriptive studies, and mixed methods studies. For each study design, the MMAT provides 5 quality criteria that must be rated with "Yes," "No," or "Can't tell." Since the calculation of an overall score from the ratings of each criterion is discouraged [36,37], we reported a separate score for each rating. Nevertheless, an overall score was reported, because it provides a general picture of study quality. Studies were not excluded based on study quality [36]. For the randomized controlled trials and nonrandomized designs, we rated the criterion "Are there complete outcome data?" as "No" when the drop-out rate was over 20% [38]. In nonrandomized designs, we rated the criterion "Are the confounders accounted for in the design and analysis?" as "No" when there was no description of additional therapy offered during the study, functional status, or cognitive status. Quality assessment was carried out by one author, and 10% of the included studies were selected at random and additionally assessed by a second author to check interrater reliability.

Data Analysis and Data Synthesis

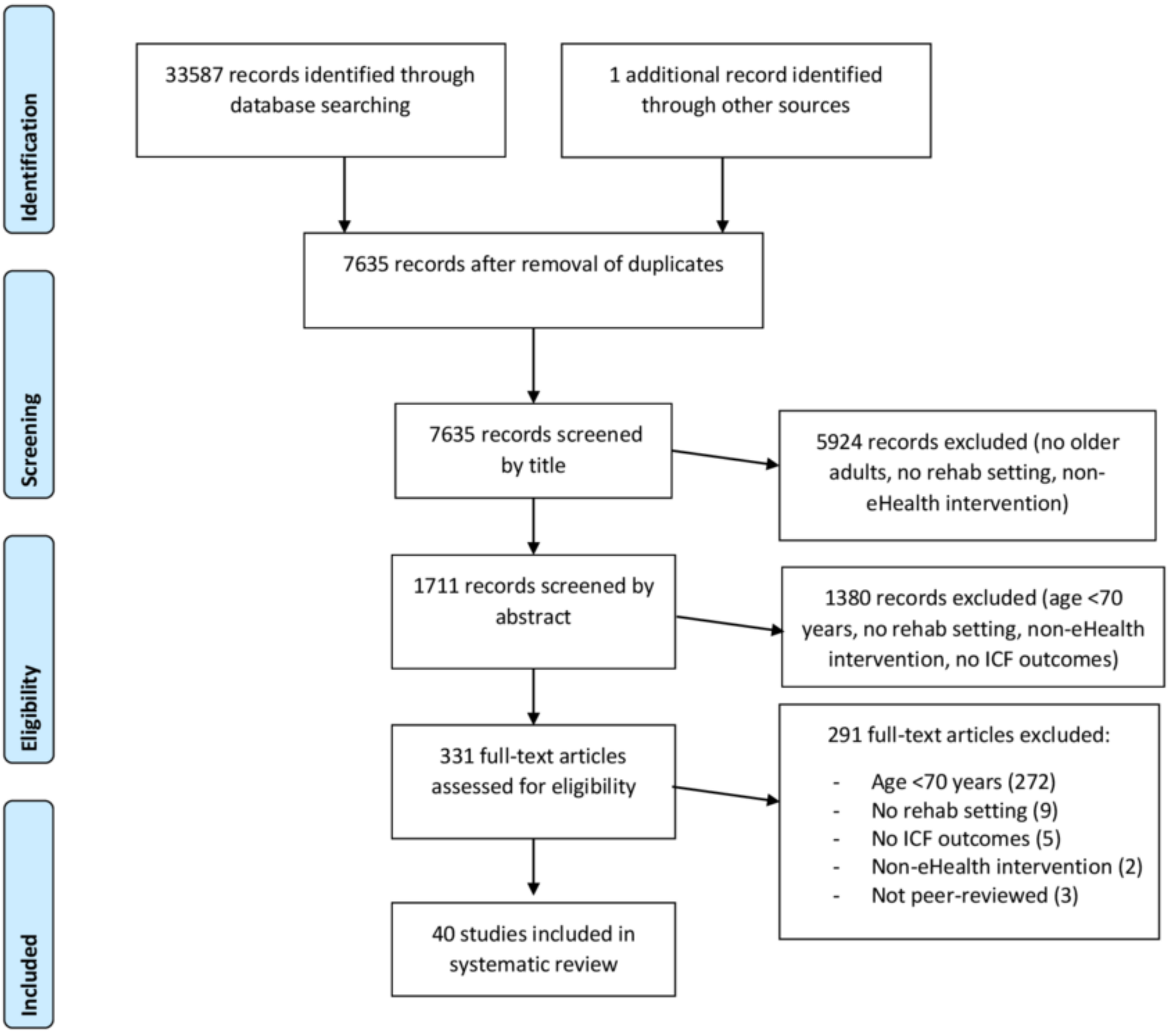
In studies that reported outcomes related to effectiveness and included a control group, a narrative synthesis was conducted using a harvest plot [39]. In the harvest plot, primary and secondary outcomes were described and color coded based on ICF domain. For each study, the bars in the harvest plot indicated the total results for the different ICF domains, and the height of the bars represented the methodical quality based on the MMAT. When a study reported multiple consistent results within the same ICF domain, the results were combined in 1 bar. If a study reported conflicting results within the same ICF domain, both results were presented. Randomized controlled trails were represented by a thick contour around bars. A meta-analysis was not feasible since the included studies were too heterogeneous with regard to population, intervention, and outcome measures.

Results

Study Selection

The search strategy identified a total of 7635 unique records. After exclusion of records based on title and abstract, 331 records remained. During full-text screening, a further 291 records were excluded, resulting in the inclusion of 40 studies in this review. Reasons for exclusion are presented in the study flowchart shown in Figure 1. In 12 cases, a third reviewer was needed to achieve consensus during the process of study selection.

Figure 1. PRISMA flow diagram of search strategy results. ICF: International Classification of Functioning, Disability, and Health.


Identification
Screening
Eligibility
Included
33587 records identified through database searching
1 additional record identified through other sources
7635 records after removal of duplicates
7635 records screened by title
5924 records excluded (no older adults, no rehab setting, non-eHealth intervention)
1711 records screened by abstract
1380 records excluded (age <70 years, no rehab setting, non-eHealth intervention, no ICF outcomes)
331 full-text articles assessed for eligibility
291 full-text articles excluded:
- Age <70 years (272)
- No rehab setting (9)
- No ICF outcomes (5)
- Non-eHealth intervention (2)
- Not peer-reviewed (3)
40 studies included in systematic review

Study Characteristics

Study characteristics are shown in Table 1. Of the 40 included studies, 18 (45%) were randomized controlled trails [40-57], 2 (5%) had a mixed methods design [58,59], 1 was a qualitative study [60], and 19 (48%) had a quantitative nonrandomized design [61-79], of which 9 studies (of 19, 47%) included a control group [53,61-68,79]. Of 40 studies, 17 studies (43%) were conducted in a hospital setting [41-44,46,50,51,55-57,62,64-66,68,71,79]. Of the 17 hospital-setting studies, 12 (71%) were conducted in a dedicated hospital-rehabilitation unit [41-44,46,50,51,55,56,64,71,79], 2 (12%) were in a hospital-stroke unit [57,68], and 1 (6%) was conducted in a geriatric day hospital [62]. Of the 40 studies, 10 (25%) were conducted in an ambulatory setting [47,48,52-54,60,69,75,76,78], 9 studies (23%) took place in a geriatric rehabilitation setting [40,45,49,58,59,61,63,70,74], 2 studies (5%) were at a tertiary rehabilitation center [60,73], 1 study (3%) was at a skilled nursing facility [77], and 2 studies (5%) did not report the setting [67,72].

Table 1. Study characteristics.

Author, year, country	Design	Diagnosis; n; setting	Age (SD); female (%)	Intervention	Use of intervention	Primary outcome domain (primary outcome measure)	Secondary outcome domain(s)
Barnason [53], 2009, United States	RCT[a]	Cardiac; n=55; Ambulatory	71.6 (5.1); 16	Video communication in combination with non-eHealth vs usual care	Daily use, subjects responded to assessment queries, were provided with strategies	Effectiveness, activities (other)	Effectiveness, participation
Backman [59], 2020, United Kingdom	Mixed methods	Orthopedic; n=30; Geriatric rehabilitation	81 (67-96); 63	Mobile apps	Providing access to discharge records during transition to home	Usability	—[b]
Bernocchi [52], 2018, Italy	RCT	Multiple diagnoses; n=146; Ambulatory	79 (6.5); 84	Video communication in combination with non-eHealth vs usual care	Weekly calls; video communication 2×/month; fall prevention program provided by therapist	Effectiveness, activities (other)	Feasibility, effectiveness, activities, participation
Bernocchi [69], 2016, Italy	Quantitative; nonrandomized	Stroke; n=15; Ambulatory	71 (11); 47	Video communication in combination with health sensors	Weekly calls with nurse; weekly video communication with physiotherapist	Feasibility (n completed, n sessions)	Effectiveness, body functions, activities
Cannell [44], 2017, Australia	RCT	Stroke; n=40; Hospital, rehabilitation unit	74 (10); 37.5	Exergames in combination with virtual reality vs usual care	1 hour/session, 5 days/week, in addition to conventional therapy	Effectiveness, activities (maintaining body position)	Effectiveness, activities
Chan [62], 2012, China	Quantitative nonrandomized	Multiple diagnoses; n=90; Geriatric Day hospital	80 (7.1); 73	Exergames vs usual care	10 min/session, 8 sessions total, in addition to conventional therapy	Feasibility (total time spent, average BS[c] and %MHR[d])	Effectiveness, activities
Cimarolli [77], 2017, United States	Quantitative; nonrandomized	Multiple diagnoses; n=237; Skilled nurse facility	76 (10.7); 59	Exergames	Recommended use: 2 sessions/week for 15 min, in addition to conventional therapy	Feasibility (time spent, predictors of intense use)	Effectiveness, external factors
Dakin [61], 2011, Australia	Quantitative; nonrandomized	Multiple diagnoses; n=34; Geriatric rehabilitation	77; 47	Health sensors vs usual care	Wore health sensor daily during admission	Effectiveness activities (ADL[e])	Effectiveness, external factors
Da-Silva [57], 2019, United Kingdom	RCT	Stroke; n=33; Hospital, stroke unit	71; 60.6	Health sensors with reminders vs health sensors without reminders	Wore health sensor for 4 weeks, health sensor vibrated to remind patients to use affected arm	Effectiveness, activities (hand and arm use)	Feasibility, adherence
Doornebosch [70], 2007, Netherlands	Quantitative; nonrandomized	Stroke; n=10; Geriatric rehabilitation	72 (53-94); 80	Robotics	20 minutes/session, 8 sessions total, in addition to conventional therapy	Personal factors (patient’s experience)	Effectiveness, body functions
Edmans [68], 2009, United Kingdom	Quantitative; nonrandomized	Stroke; n=13; Hospital, stroke unit	73; 23	Virtual reality vs usual care	1 hour/session, 5 days/week	Effectiveness, activities (other)	Effectiveness, activities
Franceschini [56], 2020, Italy	RCT	Stroke; n=48; Hospital, rehabilitation unit	72 (64.3); 45.8	Robotics vs usual care	30 minutes/session, 5 days/week over 6 weeks, in addition to conventional therapy	Effectiveness, body functions (muscle power, tone, and reflexes)	Effectiveness, (muscle power, tone, and reflexes)
Gandolfi [71], 2017, Italy	Quantitative; nonrandomized	Stroke; n=2; Hospital, rehabilitation unit	74; 100	Robotics	20 minutes/session, 5 days/week, 10 sessions total, in addition to conventional therapy	Feasibility (compliance, time to set device)	Effectiveness, body functions
Goto [65], 2017, Japan	Quantitative; nonrandomized	Orthopedic; n=20; Hospital	74 (7.5); 90	Robotics vs usual care	Every other day, in addition to conventional therapy	Effectiveness, body functions (mobility of joints)	Effectiveness, body functions

Author, year, country	Design	Diagnosis; n; setting	Age (SD); female (%)	Intervention	Use of intervention	Primary outcome domain (primary outcome measure)	Secondary outcome domain(s)
Hesse [42], 2014, Germany	RCT	Stroke; n=50; Hospital, rehabilitation unit	70 (16); 44	Robotics vs usual care	30 minutes/session, 4 days/week, in addition to conventional therapy	Effectiveness, body functions (muscle power, tone, and reflexes)	Effectiveness, body functions, activities, external factors
Hesse [72], 2010, Germany	Quantitative; nonrandomized	Stroke; n=1; Not reported	72; 0	Robotics	25 minutes/session, 5 days/week, 25 sessions in total, in addition to conventional therapy	Effectiveness, body functions (ADL)	—
Hicks [63], 2016, United States	Quantitative; nonrandomized	Cardiac; n=285; Geriatric rehabilitation	79 (48-99); 54.3	Health gateway vs usual care	Encouraged daily use, in addition to conventional therapy	Effectiveness, activities (ADL)	Effectiveness, external factors
Iosa [46], 2015, Italy	RCT	Stroke; n=4; Hospital, rehabilitation unit	71.5 (4.51); 50	Exergames in combination with virtual reality vs usual care	30 minutes/session, 3 days/week, in addition to conventional therapy	Feasibility (motivation, time spent, adverse events)	Effectiveness, body functions, activities
Karner [55], 2019, Germany	RCT	Stroke; n=56.4%; Hospital, rehabilitation unit	73,7 (7.33); 56.4	Robotics vs book reading	30 minutes/session 3 days/week over 3 weeks	Effectiveness, body functions (visual)	—
Koneva [67], 2018, Russia	Quantitative; nonrandomized	Stroke; n=40; Not reported	84 (1.2); 30	Virtual reality vs usual care	Task-specific training	Effectiveness, body functions (neurological)	Effectiveness, body functions, activities, participation
Laver [50], 2012, Australia	RCT	Multiple diagnoses; n=44; Hospital, rehabilitation unit	84.9 (4.5); 80	Exergames vs usual care	25 minutes/session, 5 days/week for duration of stay	Effectiveness, activities (mobility)	Effectiveness, body functions, activities, participation
Levinger [64], 2016, Italy	Quantitative; nonrandomized	Orthopedic; n=4; Hospital, rehabilitation unit	70; 76	Exergames vs usual care	2 sessions/week, in addition to conventional therapy	Effectiveness, activities (mobility)	Effectiveness, body functions, activities, participation
Li [54], 2020, Hong Kong	RCT	Orthopedic; n=31; Ambulatory	79,3 (9.1); 80.6	Mobile apps vs usual care	Use of app based on rehabilitation goals, in addition to conventional therapy	Effectiveness, activities (mobility)	Effectiveness, feasibility, body functions, activities,
Ling [58], 2017, Netherlands	Mixed methods	Orthopedic; n=7; Geriatric rehabilitation	70 (8); 71	Exergames	30 minutes/session, in addition to conventional therapy	Usability (ease of use)	—
Marschollek [75], 2014, Germany	Quantitative; nonrandomized	Orthopedic; n=14; Ambulatory	83.5 (71-90)	Health sensors	Sensors placed at home for monitoring ADL	Feasibility (installation time, downtimes)	Acceptability
Oesch [49], 2017, Switzerland	RCT	Multiple diagnoses; n=54; Geriatric rehabilitation	74 (67-79); 45	Exergames vs self-regulated exercises	30 minutes/session, twice a day	Effectiveness (personal factors)	Effectiveness personal factors, activities
Peel [40], 2016, Australia	RCT	Multiple diagnoses; n=270; Geriatric rehabilitation	81 (8); 58	Health sensors with goal-setting vs health sensors without goal-setting	Daily feedback and goal-setting by therapists, in addition to conventional therapy	Effectiveness, activities (mobility)	Effectiveness, activities, participation, external factors
Peel [78], 2011, Australia	Quantitative; nonrandomized	Multiple diagnoses; n=0; Ambulatory	—	Video communication	All communication conducted through intervention	Feasibility	—
Piqueras [47], 2013, Spain	RCT	Orthopedic; n=142; Ambulatory	73.3 (6.5); 72.4	Video communication in combination with health sensors vs usual care	1 hour/session over 10 days	Effectiveness, body functions (mobility of joints)	Effectiveness, body functions, activities

Author, year, country	Design	Diagnosis; n; setting	Age (SD); female (%)	Intervention	Use of intervention	Primary outcome domain (primary outcome measure)	Secondary outcome domain(s)
Pol [48], 2019, Netherlands	RCT	Orthopedic; n=240; Ambulatory	83 (6.9); 79.6	Health sensors in combination with non-eHealth intervention vs non-eHealth intervention vs usual care	Sensors placed at home for monitoring ADL, 4 home visits and 4 telephone consultations	Effectiveness, activities (other)	Effectiveness, participation
Sampson [73], 2012, New Zealand	Quantitative; nonrandomized	Stroke; n=1; Rehabilitation center	76; 100	Robotics in combination with virtual reality	45 minutes/session, 4 sessions/week over 6 weeks, in addition to conventional therapy	Effectiveness, body functions (muscle power, tone, and reflexes)	Effectiveness body functions
Schoone [45], 2011, Netherlands	RCT	Stroke; n=24; Geriatric rehabilitation	71.3 (8.2); 33	Robotics	10-30 minutes/sessions, 3 sessions/week over 6 weeks, in addition to conventional therapy	Effectiveness, body functions, activities (hand and arm use)	Effectiveness participation, external factors
Schwickert [74], 2011, Germany	Quantitative; nonrandomized	Orthopedic; n=8; Geriatric rehabilitation	79.5; 50	Robotics, virtual reality	30-45 minutes/session, 2-3 sessions/week for 2-4 weeks, in addition to conventional therapy	Feasibility (adherence, satisfaction)	Effectiveness, body functions, activities, participation
Takano [79], 2020, Japan	Quantitative; nonrandomized	Orthopedic; n=27; Hospital, rehabilitation unit	81 (6.3); 89	Robotics in combination with exergames	20 min/session 6 sessions/week for 2 weeks in addition to conventional therapy	Effectiveness activities (mobility)	Effectiveness, activities,
Taveggia [43], 2016, Italy	RCT	Stroke; n=28; Hospital, rehabilitation unit	72 (6); 39	Robotics vs usual care	30 minutes/session, 5 sessions/week over 5 weeks, in addition to conventional therapy	Effectiveness, activities (mobility)	Effectiveness, activities, participation
Tousignant [76], 2006, Canada	Quantitative; nonrandomized	Multiple diagnoses; n=4; Ambulatory	70,75; 50	Video communication	1 hour/session, 3 sessions/week over 4 weeks	Effectiveness, activities (ADL)	Effectiveness, body functions, activities
Van den Berg [51], 2015, Australia	RCT	Multiple diagnoses; n=58; Hospital, rehabilitation unit	80 (12); 62	Exergames vs usual care	1 hour/session, 5 session/week, in addition to conventional therapy	Effectiveness, activities (mobility)	Usability; Effectiveness, activities, participation
Vanoglio [41], 2017, Italy	RCT	Stroke; n= 30; Hospital, rehabilitation unit	71 (12); 53	Robotics vs usual care	40 minutes/session, 5 sessions/week over 6 weeks	Feasibility (n completed, adverse events, difficulty)	Effectiveness, body functions, external factors
White [60], 2015, Australia	Qualitative	Stroke; N=12; Rehabilitation center, ambulatory	73 (53-83); 33	Mobile apps	Therapist installed apps; patients encouraged to explore iPad	Usability	—
Yoshikawa [66], 2018, Japan	Quantitative; nonrandomized	Orthopedic; n=19; Hospital	76 (6.85); 81	Robotics vs usual care	14 minutes/session, 12-14 session in 4 weeks, in addition to conventional therapy	Effectiveness, activities (mobility)	Effectiveness, body functions

[a] RCT: randomized controlled trial.

[b] Not available.

[c] BS: Borg Perceived Exertion Scale.

[d] %MHR: maximum heart rate.

[e] ADL: activities of daily living.

Of 40 studies, 17 (43%) included participants who were diagnosed with stroke [41-46,55-57,60,67-73], 10 (25%) included participants with multiple diagnoses [40,49-52,59,61,62,76-78], 11 (28%) included participants with orthopedic problems [47,48,54,58,59,64-66,74,75,79], and 2 studies (5%) included participants with cardiac-related diagnoses [53,63]. Across all studies, the included sample size ranged from 1 to 285 participants.

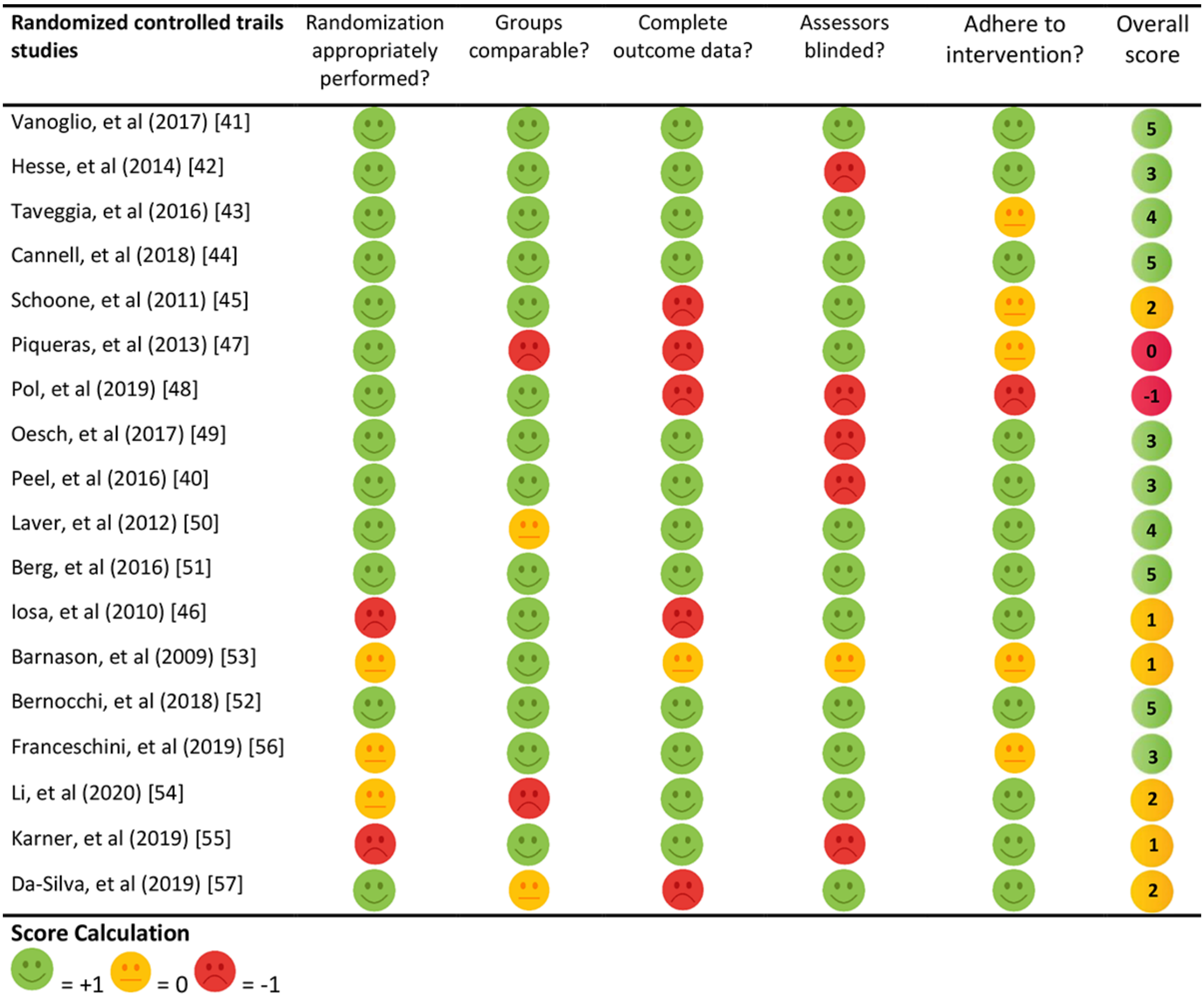
Various types of eHealth interventions were used. Of 40 studies, 11 studies (28%) delivered the intervention via robotics [41-43,45,55,56,65,66,70-72], 2 studies (5%) combined robotics with virtual reality [73,74], and 1 study (3%) combined robotics with exergames [79]. Additionally, 9 studies (of 40, 23%) investigated exergames [44,46,49-51,58,62,64,77], of which 2 (of 9, 22%) combined exergames with virtual reality [44,46] and 1 (of 9, 11%) combined exergames with health sensors [51]. Of 40 studies, 2 (5%) examined video communication [76,78], 3 (8%) combined video communication with health sensors [47,53,69], and 1 (3%) combined video communication with a non-eHealth intervention [52]. Of 40 studies, health sensors were used in 6 studies (15%) [40,48,57,61,63,75], including 1 (of 6, 17%) in combination with a health gateway [63] and 1 (of 6, 17%) in combination with a non-eHealth intervention [48]. Of 40 studies, 3 studies (8%) investigated mobile apps [54,59,60], and 2 studies (5%) examined virtual reality [67,68].

Outcome measures related to effectiveness were reported in 24 of 40 studies (60%) [40,42-45,47-50,53,55,56,61,63-68,70,72,73,76,79], and 10 of 40 studies (25%) included outcome measures related to effectiveness and feasibility [41,46,52,54,57,62,69,71,74,77]. Of 40 studies, 2 studies (5%) included outcomes related to usability [58,60], 2 studies (5%) included outcomes related only to feasibility [75,78], 1 study (3%) included outcomes related to effectiveness and usability [51], and 1 study (3%) included outcomes related to feasibility and usability [59]. A detailed description of all included studies can be found in Multimedia Appendix 3.

Study Quality

Results of the quality assessment are presented in Figure 2 and Figure 3. The quality of the included studies ranged from –3 to 5 (on a scale ranging from –5 to 5). The mean overall score was 3 for randomized controlled trails, 1 for quantitative nonrandomized studies, 1 for a mixed methods studies, and 5 for a qualitative study (based on 1 study). In quantitative nonrandomized studies, the most frequent shortcoming was insufficient reporting of confounders; only 2 of 19 studies (11%) accounted for confounders in design and analysis [73,79]. The representativeness of the target population in quantitative nonrandomized studies was also often insufficient; 9 of the 19 studies (47%) reported insufficient information, lacking either adequate explanation of why certain eligible participants chose not to participate or a clear description of the target population [53,61,65,67,69,71,75,76,78]. Additionally, 6 of the 19 studies (32%) included a sample size of less than 20 [64,66,70,72-74].

Figure 2. Quality appraisal for randomized controlled trial studies.

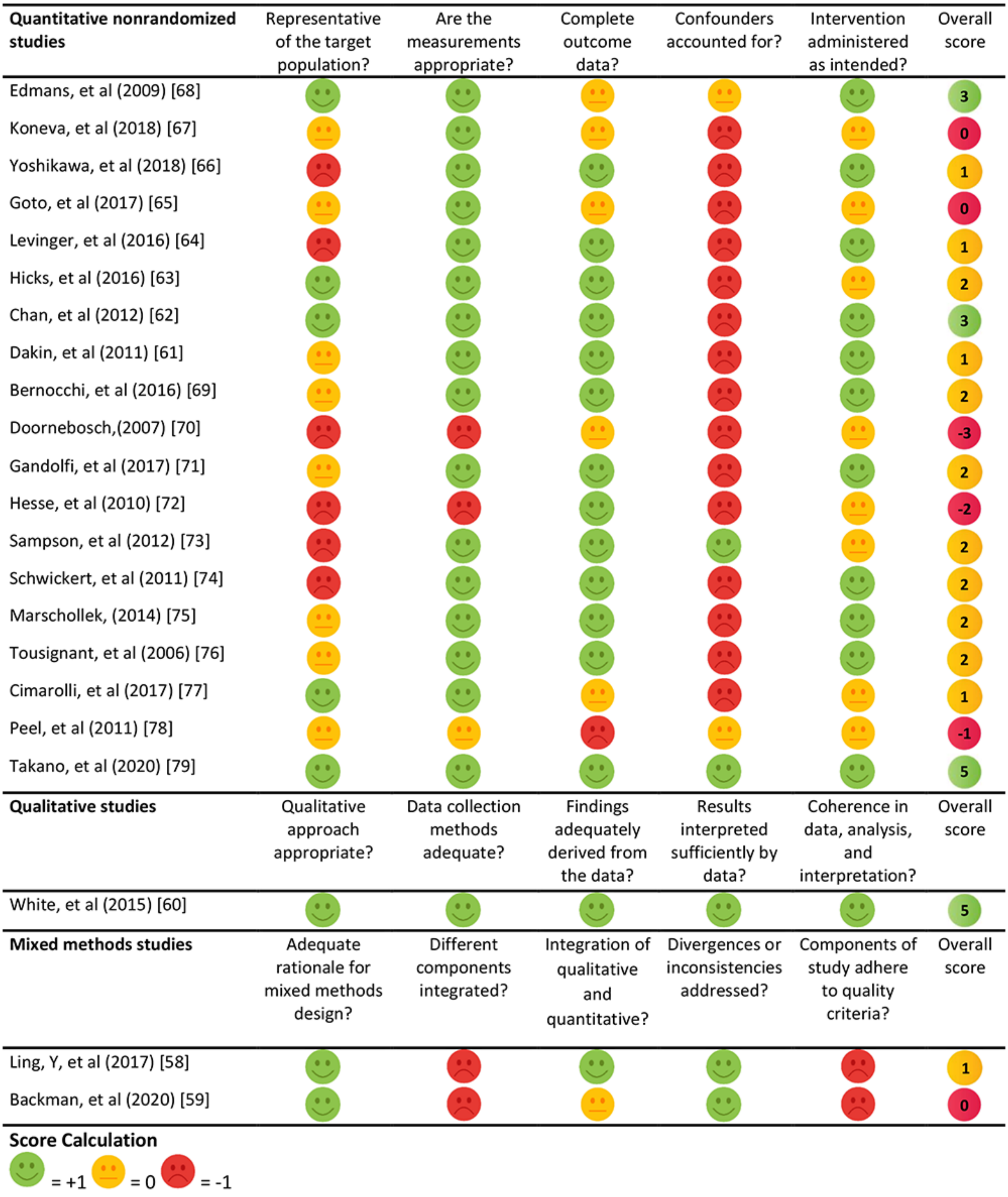


Randomized controlled trails studies	Randomization appropriately performed?	Groups comparable?	Complete outcome data?	Assessors blinded?	Adhere to intervention?	Overall score
Vanoglio, et al (2017) [41]	🟢	🟢	🟢	🟢	🟢	5
Hesse, et al (2014) [42]	🟢	🟢	🟢	🔴	🟢	3
Taveggia, et al (2016) [43]	🟢	🟢	🟢	🟢	🟡	4
Cannell, et al (2018) [44]	🟢	🟢	🟢	🟢	🟢	5
Schoone, et al (2011) [45]	🟢	🟢	🔴	🟢	🟡	2
Piqueras, et al (2013) [47]	🟢	🔴	🔴	🟢	🟡	0
Pol, et al (2019) [48]	🟢	🟢	🔴	🔴	🔴	-1
Oesch, et al (2017) [49]	🟢	🟢	🟢	🔴	🟢	3
Peel, et al (2016) [40]	🟢	🟢	🟢	🔴	🟢	3
Laver, et al (2012) [50]	🟢	🟡	🟢	🟢	🟢	4
Berg, et al (2016) [51]	🟢	🟢	🟢	🟢	🟢	5
Iosa, et al (2010) [46]	🔴	🟢	🔴	🟢	🟢	1
Barnason, et al (2009) [53]	🟡	🟢	🟡	🟡	🟡	1
Bernocchi, et al (2018) [52]	🟢	🟢	🟢	🟢	🟢	5
Franceschini, et al (2019) [56]	🟡	🟢	🟢	🟢	🟡	3
Li, et al (2020) [54]	🟡	🔴	🟢	🟢	🟢	2
Karner, et al (2019) [55]	🔴	🟢	🟢	🔴	🟢	1
Da-Silva, et al (2019) [57]	🟢	🟡	🔴	🟢	🟢	2

Score Calculation

🟢 = +1 🟡 = 0 🔴 = -1

Figure 3. Quality appraisal for quantitative nonrandomized, qualitative, and mixed methods studies.



Quantitative nonrandomized studies	Representative of the target population?	Are the measurements appropriate?	Complete outcome data?	Confounders accounted for?	Intervention administered as intended?	Overall score
Edmans, et al (2009) [68]	🙂	🙂	😐	😐	🙂	3
Koneva, et al (2018) [67]	😐	🙂	😐	☹	😐	0
Yoshikawa, et al (2018) [66]	☹	🙂	🙂	☹	🙂	1
Goto, et al (2017) [65]	😐	🙂	😐	☹	😐	0
Levinger, et al (2016) [64]	☹	🙂	🙂	☹	🙂	1
Hicks, et al (2016) [63]	🙂	🙂	🙂	☹	😐	2
Chan, et al (2012) [62]	🙂	🙂	🙂	☹	🙂	3
Dakin, et al (2011) [61]	😐	🙂	🙂	☹	🙂	1
Bernocchi, et al (2016) [69]	😐	🙂	🙂	☹	🙂	2
Doornebosch,(2007) [70]	☹	☹	😐	☹	😐	-3
Gandolfi, et al (2017) [71]	😐	🙂	🙂	☹	🙂	2
Hesse, et al (2010) [72]	☹	☹	🙂	☹	😐	-2
Sampson, et al (2012) [73]	☹	🙂	🙂	🙂	😐	2
Schwickert, et al (2011) [74]	☹	🙂	🙂	☹	🙂	2
Marschollek, (2014) [75]	😐	🙂	🙂	☹	🙂	2
Tousignant, et al (2006) [76]	😐	🙂	🙂	☹	🙂	2
Cimarolli, et al (2017) [77]	🙂	🙂	😐	☹	😐	1
Peel, et al (2011) [78]	😐	😐	☹	😐	😐	-1
Takano, et al (2020) [79]	🙂	🙂	🙂	🙂	🙂	5

Qualitative studies	Qualitative approach appropriate?	Data collection methods adequate?	Findings adequately derived from the data?	Results interpreted sufficiently by data?	Coherence in data, analysis, and interpretation?	Overall score
White, et al (2015) [60]	🙂	🙂	🙂	🙂	🙂	5

Mixed methods studies	Adequate rationale for mixed methods design?	Different components integrated?	Integration of qualitative and quantitative?	Divergences or inconsistencies addressed?	Components of study adhere to quality criteria?	Overall score
Ling, Y, et al (2017) [58]	🙂	☹	🙂	🙂	☹	1
Backman, et al (2020) [59]	🙂	☹	😐	🙂	☹	0

Score Calculation

🙂 = +1 😐 = 0 ☹ = -1

Effectiveness

Main Results for Effectiveness

Across all studies with a control group (n=27; 27/40, 68%), 73 different outcome measures were reported that were related to effectiveness, including 16 (22%) within the ICF domain "body functions," 40 (55%) in the domain "activities," 11 (15%) in the domain "participation," 4 (5%) in the domain "external factors," and 2 (3%) in the domain "personal factors" (Figure 4). In 15 studies (of 27, 56%), eHealth interventions were found to be at least as effective as non-eHealth interventions when focusing on the primary outcome measure, and 11 studies (of 27, 41%) reported eHealth interventions to be more effective than non-eHealth interventions. Of 27 studies, 1 study (4%) reported beneficial outcomes in favor of the non-eHealth interventions. Results for each ICF domain are described in detail below. A harvest plot illustrating the evidence regarding effectiveness is presented in Figure 5.

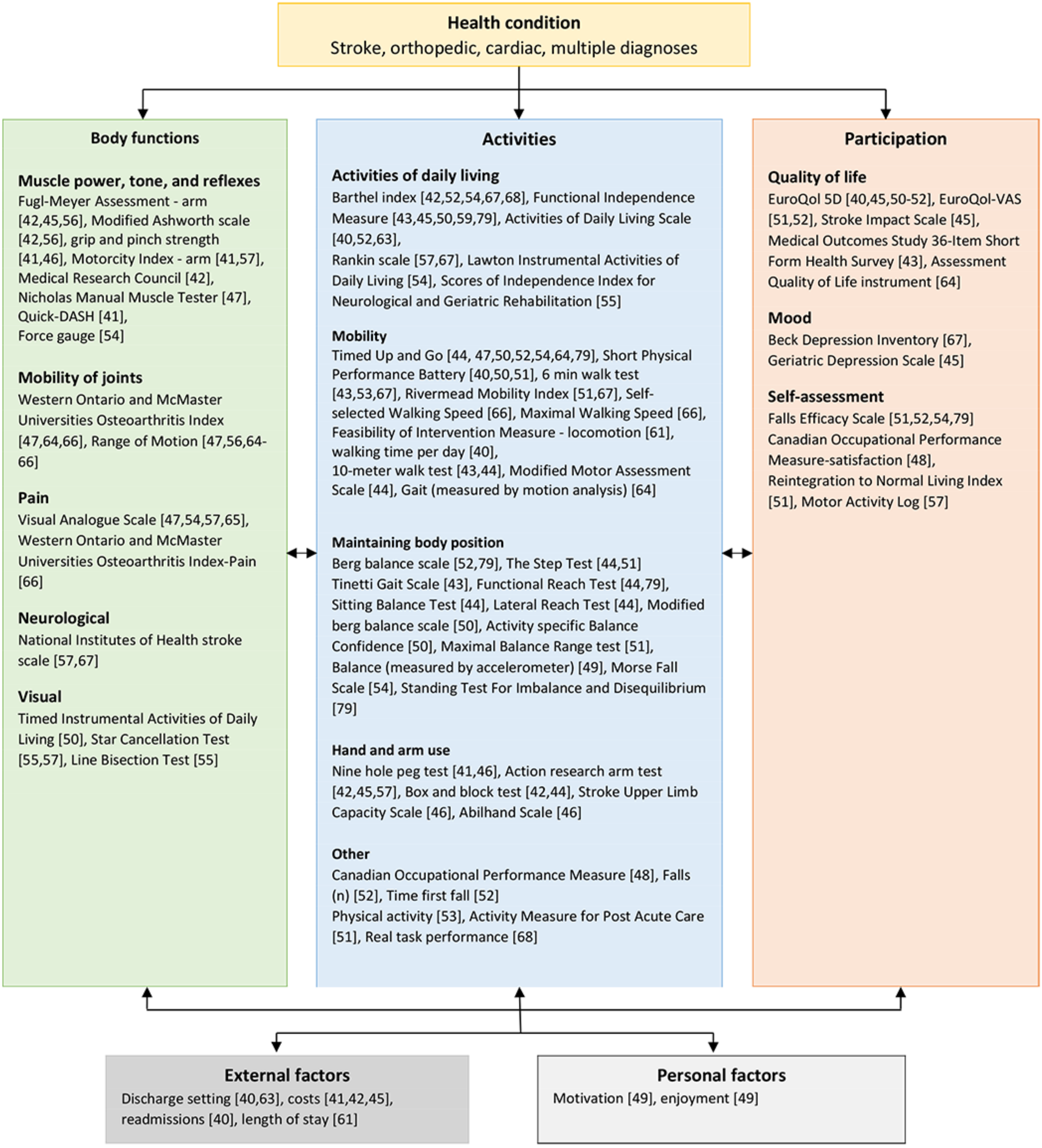
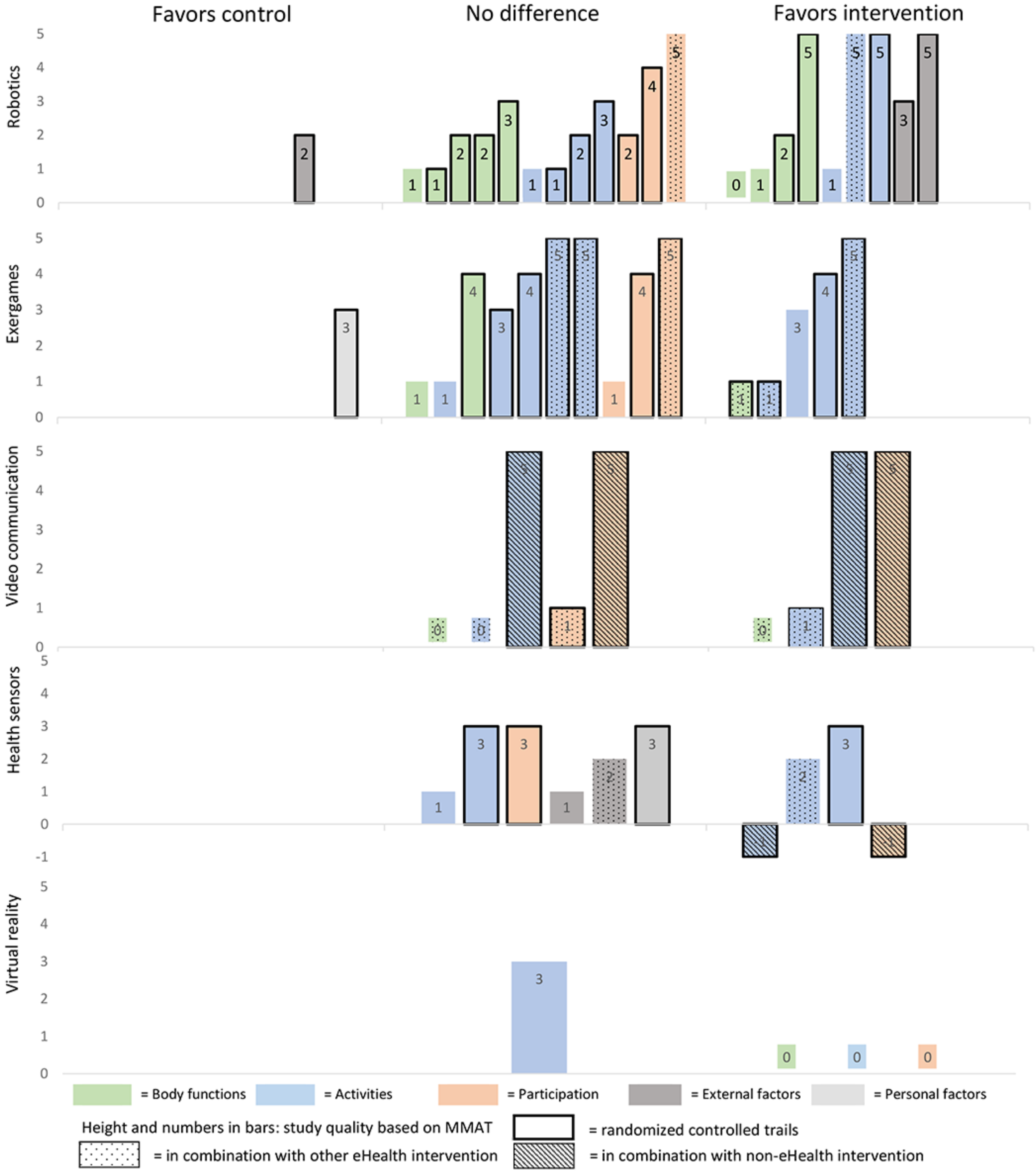


Figure 4. Outcome measures classified by the International Classification of Functioning, Disability, and Health model.

Health condition
Stroke, orthopedic, cardiac, multiple diagnoses
Body functions
Muscle power, tone, and reflexes
Fugl-Meyer Assessment - arm [42,45,56], Modified Ashworth scale [42,56], grip and pinch strength [41,46], Motorcity Index - arm [41,57], Medical Research Council [42], Nicholas Manual Muscle Tester [47], Quick-DASH [41], Force gauge [54]
Mobility of joints
Western Ontario and McMaster Universities Osteoarthritis Index [47,64,66], Range of Motion [47,56,64-66]
Pain
Visual Analogue Scale [47,54,57,65], Western Ontario and McMaster Universities Osteoarthritis Index-Pain [66]
Neurological
National Institutes of Health stroke scale [57,67]
Visual
Timed Instrumental Activities of Daily Living [50], Star Cancellation Test [55,57], Line Bisection Test [55]
Activities
Activities of daily living
Barthel index [42,52,54,67,68], Functional Independence Measure [43,45,50,59,79], Activities of Daily Living Scale [40,52,63],
Rankin scale [57,67], Lawton Instrumental Activities of Daily Living [54], Scores of Independence Index for Neurological and Geriatric Rehabilitation [55]
Mobility
Timed Up and Go [44, 47,50,52,54,64,79], Short Physical Performance Battery [40,50,51], 6 min walk test [43,53,67], Rivermead Mobility Index [51,67], Self-selected Walking Speed [66], Maximal Walking Speed [66], Feasibility of Intervention Measure - locomotion [61], walking time per day [40],
10-meter walk test [43,44], Modified Motor Assessment Scale [44], Gait (measured by motion analysis) [64]
Maintaining body position
Berg balance scale [52,79], The Step Test [44,51]
Tinetti Gait Scale [43], Functional Reach Test [44,79], Sitting Balance Test [44], Lateral Reach Test [44], Modified berg balance scale [50], Activity specific Balance Confidence [50], Maximal Balance Range test [51], Balance (measured by accelerometer) [49], Morse Fall Scale [54], Standing Test For Imbalance and Disequilibrium [79]
Hand and arm use
Nine hole peg test [41,46], Action research arm test [42,45,57], Box and block test [42,44], Stroke Upper Limb Capacity Scale [46], Abilhand Scale [46]
Other
Canadian Occupational Performance Measure [48], Falls (n) [52], Time first fall [52]
Physical activity [53], Activity Measure for Post Acute Care [51], Real task performance [68]
Participation
Quality of life
EuroQol 5D [40,45,50-52], EuroQol-VAS [51,52], Stroke Impact Scale [45], Medical Outcomes Study 36-Item Short Form Health Survey [43], Assessment Quality of Life instrument [64]
Mood
Beck Depression Inventory [67], Geriatric Depression Scale [45]
Self-assessment
Falls Efficacy Scale [51,52,54,79]
Canadian Occupational Performance Measure-satisfaction [48],
Reintegration to Normal Living Index [51], Motor Activity Log [57]
External factors
Discharge setting [40,63], costs [41,42,45], readmissions [40], length of stay [61]
Personal factors
Motivation [49], enjoyment [49]

Figure 5. Harvest plot: effectiveness of eHealth interventions. MMAT: Mixed Methods Appraisal Tool.



Body Functions

Of 40 total studies, 14 studies (35%) included 16 outcomes related to body functions [41,42,45-47,50,54-57,64-67]. Of these 14 studies, 9 studies (64%) found, in 7 outcome measures, significant improvements in favor of the intervention group (Figure 5) [41,46,47,54-56,65-67]. Of 14 studies, 4 studies (29%) reported improved muscle power through robotics [56,65], exergames [46], or mobile apps [54]. Of 14 studies, 4 studies (29%) found that the addition of robotics [56,65,66] or video communication in combination with health sensors [47] improved the mobility of joints when compared with physical therapy alone. Another 2 studies (of 14, 14%) reported that the use of robotics could decrease pain when compared with conventional physiotherapy [65,66]. Koneva and colleagues [67] reported that the use of virtual reality improved neurological status, as measured by the National Institutes of Health stroke scale, when compared with usual care (5.2±0.4 vs 6.3±0.5; *P*<.001).

Activities

Of all 40 included studies, 25 studies (63%) reported 40 outcomes related to activities [40-55,57,61-64,66-68,79], and 13 studies (33%) found, in 17 outcomes, a significant outcome in favor of the intervention group [40,41,46,48,50-53,62,63,66,67,79]. Of 40 studies, 5 studies (13%) demonstrated that eHealth was effective in improving activities of daily living when the intervention was delivered via video communication in combination with health sensors and a non-eHealth intervention [52] or when the intervention was delivered via health sensors in combination with health gateways [63], exergames [62], robotics [79], or virtual reality [67]. In these studies, eHealth was compared with usual care [52,67], physiotherapy [62,79], or no intervention [63]. Another 6 studies (of 40, 15%) found that eHealth could contribute to improved mobility through the use of robotics [52,79], exergames [50], virtual reality [67], video communication in combination with health sensors [52], or health sensors in combination with goal setting [40]. These interventions were compared with physiotherapy [50,66,79], usual care [52,67], or health sensors without goal setting [40]. Of 40 studies, 4 studies (10%) reported improvements in balance when the intervention was delivered via robotics [79], exergames [50], exergames in combination with health sensors [51], or video communication in combination with health sensors [52], when compared with physiotherapy [50,51,79] or usual care [52]. Another 2 studies (of 40, 5%) reported that either robotics [41] or exergames in combination with health sensors [46] could improve hand and arm function when compared with physiotherapy [41] or no intervention [46]. Pol and colleagues [48] found that patient-reported daily functioning significantly improved with the use of health sensors in combination with cognitive behavioral treatment, compared with cognitive behavioral treatment alone, reporting a difference of 1.17 (95% CI 0.47-1.87; *P*<.001). Bernocchi and colleagues [52] reported that the use of video communication in combination with health sensors and a non-eHealth intervention was effective in preventing falls in patients who were at high risk of falling, when compared with usual care (29 falls vs 56 falls; *P*<.001). Of 40 studies, 1 study (3%) demonstrated that the use of video communication in combination with health sensors improved physical activity when compared with usual care [53].

Participation

Of 40 studies, 12 studies (30%) included 11 outcome measures within the participation domain [40,43,45,48,50-53,57,64,67,79]. Of these 12 studies, 3 studies (27%) reported a significant difference in quality of life [52], mood [67], or self-assessment [48] when the intervention was delivered via the use of video communication in combination with health sensors and a non-eHealth intervention [52], virtual reality [67], or the use of health sensors in combination with a non-eHealth intervention [48]. Particularly, Bernocchi and colleagues [52] demonstrated that the use of video communication in combination with health sensors and a non-eHealth intervention significantly improved scores on the EuroQol Visual Analog Scale at 6 months, when compared with usual care (mean 63.8 vs mean 53.5; *P*<.001). Koneva and colleagues [67] reported that the use of virtual reality decreased the severity of depression as measured by the Beck Depression Inventory, when compared with usual care (mean 9.5, SD 5.52 vs mean 10.3, SD 6.03; *P*<.05). Additionally, Pol and colleagues [48] found that the use of health sensors in combination with a non-eHealth intervention significantly improved the performance satisfaction in daily functioning at 6 months, when compared with usual care, reporting a difference of 0.94 (95% CI 0.37-1.52; *P*<.001).

External Factors

Across all 40 studies, 5 studies (13%) included outcome measures related to external factors [40,42,45,61,63]. Of these 5 studies, 2 studies (40%) included robotics as interventions and found significant differences in cost, in favor of the intervention group [41,42]. Of the 5 studies, 1 study (20%) included robotics as an intervention and found a difference in favor of the control group [45]. Hesse and colleagues [42] and Vanoglio and colleagues [41] reported decreases in cost with the use of robotics in comparison with either regular arm therapy (€4.15 [US $4.92] for robotic interventions vs €10.00 [US $11.85] for regular arm therapy, for each patient per session) [42] or physiotherapy (€237 [US $280.73] for robotic intervention vs €480 [US $568.57] for physiotherapy, for each patient per 30 days) [41]. In contrast, Schoone and colleagues [45] reported an increase in total costs when compared with physiotherapy (€644.14 [US $762.99] for robotic interventions vs €423.74 [US $501.93] for physiotherapy). Across all studies, no differences were found with regard to discharge settings [40,63], readmissions [40], or lengths of stay [61].

Personal Factors

Oesch and colleagues [49] found that self-regulated exercise using instruction leaflets was superior to exergames in terms of enjoyment (effect size: 0.88, range 0.32-1.44; *P*<.001) and motivation (effect size: 0.59, range 0.05-1.14; *P*=.046).

Feasibility

Main Results for Feasibility

Of the 40 included studies, 20 studies (50%) evaluated the feasibility of the eHealth intervention used [41,46,50-52,54,57,59,60,62,64,65,69,71,72,74-78], of which 19 (of 20, 95%) concluded that the eHealth intervention was feasible when it was delivered via robotics [41,65,71,72], robotics in combination with exergames [74], exergames [50,62,64,77], exergames in combination with health sensors [46,51], video communication [76], video communication in combination with health sensors [52,69], health sensors [57], health gateways in combination with health sensors [75], or mobile apps [54,59,60]. Peel and colleagues [78] reported that the use of video communication was not feasible due to problems related to patient limitations, staff issues, and the logistics of the system.

The outcome measures applied to evaluate feasibility varied considerably among studies, and a total of 19 different outcome measures were used. Of the 20 studies that reported feasibility, 6 studies (30%) reported outcomes related to "adverse events," 7 studies (35%) reported outcomes related to "adherence," and 7 studies (35%) reported outcomes related to "exclusion rate." Another 4 studies (of 20, 20%) did not specify the outcome measure used to evaluate feasibility but used outcomes related to effectiveness to establish feasibility [54,64,65,72].

Adverse Events

None of the included studies reported serious adverse events during the study period [41,46,50,51,74,76]. However, 2 studies (of 40, 5%) reported that some participants experienced discomfort during exergames [49,50].

Adherence

Of 40 studies, adherence was reported in 7 studies (18%) [49-52,57,74], and 5 studies (13%) reported information regarding the number of completed sessions [41,50-52,69]. Of the 7 studies reporting adherence, 5 studies (71%) reported high levels of adherence, ranging from 76% [52] to 100% [74]. Of the 7 studies, 2 studies (29%) reported low adherence in patients assigned to an exergame intervention when compared with either a non-eHealth intervention [49] or use of the exergame intervention below the recommended level (<30 minutes per week) [77].

Exclusion Rate

Of 40 studies, high exclusion rates were found in 7 studies (18%). Specifically, of these 7 studies, 1 study (14%) reported an exclusion rate of 64% [47], 2 studies (29%) reported an exclusion rate of 75% [49,51], and 4 studies (57%) reported an exclusion rate over 80% [42,45,50,68]. In these latter studies, eHealth was delivered through complex eHealth interventions: robotics [42,45], exergames [50], and virtual reality [68]. The most commonly reported reasons for exclusion were cognitive impairment [45,47,49,51], physical impairment [45,49], and refusal to participate [42,47,49-51,68]. Of these 7 studies, in 2 studies (29%), the reason given for declining to participate was "no interest" in eHealth [50,51].

Usability

Main Results for Usability

Of 40 studies, outcomes related to the usability of eHealth interventions were addressed in 4 studies (10%): 2 studies (5%) evaluated the usability of exergames [51,58], and another 2 studies (5%) evaluated mobile apps [59,60]. Evaluation of usability consisted of a system usability scale [51], a survey of patients and therapists [58,59], or semistructured interviews [59,60]. Of the 4 studies that reported usability, 2 studies (50%) included outcomes related to the barrier "cognition," 4 studies (100%) included outcomes related to the aging barrier "motivation," and 1 study (25%) included outcomes related to the barrier "physical ability." None of the studies included outcomes related to the barrier "perception."

Cognition

Ling and colleagues [58] reported that some patients found exergames too complicated because of the requirement to engage in multiple activities simultaneously, and they experienced difficulties in following instructions. To tailor the exergames to older patients with cognitive impairments, the authors advised to minimize the amount of information presented on the screen, which might help older patients to perceive the information better [58]. Additionally, White and colleagues [60] reported that patients with cognitive impairments experienced difficulties in operating mobile apps and needed their partner for support.

Motivation

Van den Berg and colleagues [51] reported a mean score of 62 (SD 21), on the system usability scale (scores ranging from 0 to 100), indicating that participants were generally comfortable with exergames and that they would like to use exergames more frequently. Similar findings were reported by Ling and colleagues [58], who concluded that patients and therapists both found exergames easy to use and therapists intended to use the exergame in the future. Therapists rated the exergame as highly satisfactory for motor rehabilitation in older patients after hip surgery. Findings regarding mobile apps indicated that patients readily grasped the skills required for use and that this was a beneficial source of extrinsic motivation [59,60].

Physical Ability

Ling and colleagues [58] reported that some patients with physical disabilities had difficulties playing certain exergames that required stepping exercises because these patients were unable to maintain balance during exergames.

Discussion

Principal Findings

This review aimed to provide an overview of the effectiveness, feasibility, and usability of eHealth in geriatric rehabilitation. The review included 40 studies that applied eHealth interventions in older patients receiving geriatric rehabilitation. The majority of the included studies showed that eHealth interventions in geriatric rehabilitation are at least as effective as non-eHealth interventions. All studies that delivered eHealth in combination with another non-eHealth intervention reported positive outcomes. Most studies included outcome measures related to the ICF domain "activities." Very few studies included outcomes related to the ICF domain "participation." eHealth seems to be feasible in geriatric rehabilitation, since no serious adverse events were reported and most studies reported high levels of adherence. However, high exclusion rates were found in some studies. Results related to usability indicate that there are certain age-related barriers, such as cognition and physical ability, that lead to difficulties in using eHealth. Very few studies included outcomes related to feasibility and usability. However, these are important prerequisites to maximize the likelihood of successful implementation, and they thereby influence the effectiveness of eHealth.

Comparison With Prior Work

Our findings suggest that eHealth delivered via robotics, exergames, or health sensors is often found to be at least as effective as non-eHealth. Previous reviews that examined robotics [80], exergames [16], or health sensors [81,82] often found more beneficial results in favor of the intervention group. These reviews did not focus on older adults who were admitted for geriatric rehabilitation, and this could indicate that there are certain age-related barriers that affect the effectiveness of eHealth in older adults receiving geriatric rehabilitation. All of the included studies that delivered eHealth in combination with a non-eHealth intervention reported beneficial outcomes in favor of the intervention group. This is in line with other studies in which eHealth was delivered in combination with a non-eHealth intervention [83-85]. This indicates that eHealth is more beneficial when provided through blended care, where eHealth is delivered in combination with face-to-face treatment. This may provide a better quality of care by combining the best of the two types of interventions. This seems to especially be the case when blended care is delivered via video communication [52] or health sensors [48], since it offers the possibility to monitor and treat patients remotely.

Almost all of the studies that included outcomes related to feasibility concluded that eHealth was feasible in older adults receiving geriatric rehabilitation. None of the studies reported serious adverse events, which is in line with other reviews concerning feasibility of exergames [15,86]. The majority of the studies that included outcomes related to adherence or completed sessions reported high levels of adherence. Previous reviews that examined exergames also reported high adherence rates [86]. Some studies where eHealth was delivered via robotics or exergames reported a high exclusion rate (up to 88%). All studies with exclusion rates of ≥75% were conducted in a geriatric rehabilitation setting [45,49] or in a hospital with a dedicated rehabilitation unit [50,51]. Reasons for exclusion were mostly cognitive or physical impairments, problems that are often present in older patients receiving geriatric rehabilitation. These findings indicate that eHealth in geriatric rehabilitation is safe to use and overall adherence is expected to be high, but complex eHealth interventions such as robotics and exergames might only be feasible in a selective group of older patients receiving geriatric rehabilitation.

There is limited available evidence on the usability of eHealth interventions. The studies included in our review indicate that exergames and mobile apps are usable once older patients have been trained in their use. However, there were certain age-related barriers associated with cognitive or physical ability that led to difficulties in using eHealth. While we did not find studies that reported problems in the use of eHealth due to problems in perception, 2 of 4 studies (50%) that included usability outcome measures explicitly excluded patients with visual impairments [51,58]. This might suggest that poor usability was expected in patients with visual impairments; this is in line with findings from other studies [27]. These findings suggest that usability problems are expected in older patients receiving geriatric rehabilitation, since they often suffer from cognitive, physical, or visual impairments. eHealth should be tailored to these specific age-related barriers to maximize the probability of successful use and implementation [22,27]. Furthermore, most studies did not incorporate clear usability endpoints, and the evaluation of usability varied considerably among studies. The lack of using clear endpoints or reliable and validated questionnaires combined with task metrics (preferably, task completion) to evaluate usability hampers the ability to pinpoint usability issues and prevents comparisons across different eHealth types [25,87].

Strengths and Limitations

The first strength of this review is the extensive search strategy that covered a broad range of search databases and included all types of research designs. Another strength of this review is the categorization of outcome measures based on the ICF model, providing a clear overview of different types of outcome domains evaluated in the included studies. Nonetheless, several limitations of this systematic review should be noted. While this review provides a broad overview of the literature on 3 different concepts, our study design led to a vast variety of different outcome measures related to effectiveness. The inclusion of various outcomes measures, in combination with various eHealth interventions and diagnoses, limited our ability to draw definitive conclusions. Since a meta-analysis was not feasible, we were unable to report an effect size and publication bias. We instead provided an overview of the effectiveness of eHealth interventions using a harvest plot. Lastly, while we used a separate search string that included keywords related to usability, we only found 4 studies that included outcomes on usability. A possible explanation might be that we did not include specific Computer Science search databases, which might include more studies that are related to usability [88]. Furthermore, despite the massive growth in eHealth studies, only a small portion publish their usability results [89].

Conclusions

In conclusion, eHealth can improve rehabilitation outcomes in older adults receiving geriatric rehabilitation. Based on our findings, comparisons to literature, and the strengths and limitations of our review, our main results and recommendations for further research and the use of eHealth in clinical practice are (1) keep it simple, (2) include evidence on usability, (3) focus on participation, and (4) ensure consensus. First, simple interventions have the most potential to improve rehabilitation outcomes in older adults receiving geriatric rehabilitation, especially, when they are provided as blended care. Additionally, simple eHealth interventions have a higher chance of feasibility in older patients receiving geriatric rehabilitation who often suffer from cognitive or physical impairments. Second, scarce evidence on the usability of eHealth might hamper the implementation of eHealth in older patients receiving geriatric rehabilitation and could negatively influence effectiveness and feasibility. Further research on this topic with clear endpoints is needed. Health care professionals need to be aware of the usability of eHealth interventions they are providing. Third, participation is a key concept in geriatric rehabilitation and plays an important role in enabling older patients to continue living as independently as possible. Future research on eHealth interventions should consider including outcome measures related to participation. Fourth, current evidence on the use and evaluation of eHealth in geriatric rehabilitation is diverse, making it hard to compare outcomes and draw evident conclusions. Consensus on the use and evaluation of eHealth is needed for further development and implementation of eHealth in geriatric rehabilitation.

Authors' Contributions

JJMK screened titles of the identified studies. JJMK and AV screened the abstracts of all potentially relevant studies and obtained and reviewed the full texts. Disagreements between JJMK and AV were discussed until a consensus was reached. If a disagreement could not be resolved, EFvDvI was consulted. JJMK extracted the data. AV also extracted a subset of the data (10% of included studies) to check interrater reliability. JJMK performed quality assessment, and 10% of the included studies were selected at random and additionally assessed by AV to check interrater reliability. In 12 cases, a third reviewer, EFvDvI, was needed to achieve consensus during the process of study selection.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA checklist.
[DOC File , 65 KB - jmir_v23i8e24015_app1.doc]

Multimedia Appendix 2
Keyword strings used for searching databases.
[PDF File (Adobe PDF File), 439 KB - jmir_v23i8e24015_app2.pdf]

Multimedia Appendix 3
Detailed description of all included studies.
[XLSX File (Microsoft Excel File), 2794 KB - jmir_v23i8e24015_app3.xlsx]

References

1. United Nations. Department of Economic and Social affairs, Population division. World Population Ageing. 2019. URL: https://www.un.org/en/development/desa/population/publications/pdf/ageing/WorldPopulationAgeing2019-Report.pdf [accessed 2020-11-06]
2. Salive ME. Multimorbidity in older adults. Epidemiol Rev 2013 Jan 31;35(1):75-83. [doi: 10.1093/epirev/mxs009] [Medline: 23372025]
3. Inouye S, Studenski S, Tinetti M, Kuchel G. Geriatric syndromes: clinical, research, and policy implications of a core geriatric concept. J Am Geriatr Soc 2007 May;55(5):780-791 [FREE Full text] [doi: 10.1111/j.1532-5415.2007.01156.x] [Medline: 17493201]
4. Covinsky KE, Pierluissi E, Johnston CB. Hospitalization-associated disability: "She was probably able to ambulate, but I'm not sure". JAMA 2011 Oct 26;306(16):1782-1793. [doi: 10.1001/jama.2011.1556] [Medline: 22028354]
5. Bachmann S, Finger C, Huss A, Egger M, Stuck A, Clough-Gorr K. Inpatient rehabilitation specifically designed for geriatric patients: systematic review and meta-analysis of randomised controlled trials. BMJ 2010 Apr 20;340:c1718 [FREE Full text] [doi: 10.1136/bmj.c1718] [Medline: 20406866]
6. Holstege MS, Caljouw MA, Zekveld IG, van Balen R, de Groot AJ, van Haastregt JC, et al. Successful Geriatric Rehabilitation: Effects on Patients' Outcome of a National Program to Improve Quality of Care, the SINGER Study. J Am Med Dir Assoc 2017 May 01;18(5):383-387. [doi: 10.1016/j.jamda.2016.10.011] [Medline: 27939318]
7. Lettow B, Wouters M, Sinnige J. E-Health, wat is dat? Nictiz. URL: https://www.nictiz.nl/wp-content/uploads/E-health-Wat-is-dat.pdf [accessed 2020-10-07]
8. Glasgow RE, Phillips SM, Sanchez MA. Implementation science approaches for integrating eHealth research into practice and policy. Int J Med Inform 2014 Jul;83(7):e1-11. [doi: 10.1016/j.ijmedinf.2013.07.002] [Medline: 23910896]
9. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci 2016 Oct 26;11(1):146. [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
10. Ossebaard HC, De Bruijn ACP, van Gemert-Pijnen JEWC, Geertsma RE. Risks related to the use of eHealth technologies: An exploratory study. National Institute for Public Health and the Environment. URL: https://www.rivm.nl/bibliotheek/rapporten/360127001.pdf [accessed 2020-10-06]
11. Bonten TN, Rauwerdink A, Wyatt JC, Kasteleyn MJ, Witkamp L, Riper H, EHealth Evaluation Research Group. Online Guide for Electronic Health Evaluation Approaches: Systematic Scoping Review and Concept Mapping Study. J Med Internet Res 2020 Aug 12;22(8):e17774 [FREE Full text] [doi: 10.2196/17774] [Medline: 32784173]
12. Enam A, Torres-Bonilla J, Eriksson H. Evidence-Based Evaluation of eHealth Interventions: Systematic Literature Review. J Med Internet Res 2018 Nov 23;20(11):e10971 [FREE Full text] [doi: 10.2196/10971] [Medline: 30470678]

13. Broens THF, Huis in't Veld RMHA, Vollenbroek-Hutten MMR, Hermens HJ, van Halteren AT, Nieuwenhuis LJM. Determinants of successful telemedicine implementations: a literature study. J Telemed Telecare 2007 Jun;13(6):303-309. [doi: 10.1258/135763307781644951] [Medline: 17785027]
14. Muellmann S, Forberger S, Möllers T, Bröring E, Zeeb H, Pischke CR. Effectiveness of eHealth interventions for the promotion of physical activity in older adults: A systematic review. Prev Med 2018 Mar;108:93-110. [doi: 10.1016/j.ypmed.2017.12.026] [Medline: 29289643]
15. Skjæret N, Nawaz A, Morat T, Schoene D, Helbostad JL, 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]
16. Zeng N, Pope Z, Lee J, Gao Z. A systematic review of active video games on rehabilitative outcomes among older patients. Journal of Sport and Health Science 2017 Mar;6(1):33-43 [FREE Full text] [doi: 10.1016/j.jshs.2016.12.002]
17. Reis E, Postolache G, Teixeira L, Arriaga P, Lima ML, Postolache O. Exergames for motor rehabilitation in older adults: an umbrella review. Physical Therapy Reviews 2019 Jul 19;24(3-4):84-99. [doi: 10.1080/10833196.2019.1639012]
18. Orsmond GI, Cohn ES. The Distinctive Features of a Feasibility Study: Objectives and Guiding Questions. OTJR (Thorofare N J) 2015 Jul 06;35(3):169-177. [doi: 10.1177/1539449215578649] [Medline: 26594739]
19. 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. [doi: 10.2196/mhealth.5127] [Medline: 27282195]
20. 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]
21. Kingston J. Conducting feasibility studies for knowledge based systems. Knowledge-Based Systems 2004 May;17(2-4):157-164. [doi: 10.1016/j.knosys.2004.03.011]
22. Hawley-Hague H, Boulton E, Hall A, Pfeiffer K, Todd C. Older adults' perceptions of technologies aimed at falls prevention, detection or monitoring: a systematic review. Int J Med Inform 2014 Jun;83(6):416-426 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.03.002] [Medline: 24798946]
23. Bevan N, Carter J, Earthy J, Geis T, Harker S. New ISO standards for usability, usability reports and usability measures. 2016 Jul 17 Presented at: Paper presented at: International conference on human-computer interaction; 2016; Toronto, Canada. [doi: 10.1007/978-3-319-39510-4_25]
24. Broekhuis M, Velsen L, Stal S, Weldink J, Tabak M. Why My Grandfather Finds Difficulty in using Ehealth: Differences in Usability Evaluations between Older Age Groups. In: Paper presented at: ICT4AW. 2019 Apr 23 Presented at: ICT4AW; 2019; prague czech republic p. E2019. [doi: 10.5220/0007680800480057]
25. Sousa V, Lopez KD. Towards Usable E-Health. Appl Clin Inform 2017 Dec 21;08(02):470-490. [doi: 10.4338/aci-2016-10-r-0170]
26. Alley SJ, Samra P, Rebar AL, Schoeppe S, Parkinson L, Power D, et al. A focus group study of older adults' perceptions and preferences towards web-based physical activity interventions. Inform Health Soc Care 2020 Sep 06;45(3):273-281. [doi: 10.1080/17538157.2019.1656210] [Medline: 31690152]
27. Wildenbos GA, Peute L, Jaspers M. Aging barriers influencing mobile health usability for older adults: A literature based framework (MOLD-US). Int J Med Inform 2018 Jun;114:66-75. [doi: 10.1016/j.ijmedinf.2018.03.012] [Medline: 29673606]
28. Demiris G, Finkelstein SM, Speedie SM. Considerations for the design of a Web-based clinical monitoring and educational system for elderly patients. J Am Med Inform Assoc 2001 Sep 01;8(5):468-472 [FREE Full text] [doi: 10.1136/jamia.2001.0080468] [Medline: 11522767]
29. Kraaijkamp J, van Dam van Isselt E, Chavannes N, Achterberg WP, Versluis A, Persoon A. Effectiveness of eHealth interventions in older patients who are submitted for geriatric rehabilitation: a protocol for a systematic review. PROSPERO International prospective register of systematic reviews. 2019. URL: https://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42019133192&ID=CRD42019133192 [accessed 2019-12-01]
30. Moher D, Liberati A, Tetzlaff J, Altman D, 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 [FREE Full text] [doi: 10.7326/0003-4819-151-4-200908180-00135] [Medline: 19622511]
31. van Balen R, Gordon AL, Schols JMGA, Drewes YM, Achterberg WP. What is geriatric rehabilitation and how should it be organized? A Delphi study aimed at reaching European consensus. Eur Geriatr Med 2019 Oct 08;10(6):977-987. [doi: 10.1007/s41999-019-00244-7]
32. Grund S, Gordon AL, van Balen R, Bachmann S, Cherubini A, Landi F, et al. European consensus on core principles and future priorities for geriatric rehabilitation: consensus statement. Eur Geriatr Med 2020 Apr 13;11(2):233-238. [doi: 10.1007/s41999-019-00274-1] [Medline: 32297191]
33. WHO. International classification of functioning, disability and health. WHO. International classification of functioning, disabilityhealth: World Health Organization; 2001. URL: http://apps.who.int/iris/bitstream/handle/10665/42407/9241545429.pdf;jsessionid=DD34E52C5D093BFA4F006375315882CE?sequence=1 [accessed 2020-12-19]
34. Vertias Health Innovation. Covidence systematic review software. In: Veritas Health Innovation Melbourne, VIC; 2019. URL: https://www.covidence.org/ [accessed 2019-10-10]

35. Krijgsman J, Klein WG. Ordening in de wereld van eHealth. Nictiz. 2012. URL: https://www.nictiz.nl/wp-content/uploads/2012/08/Whitepaper-Ordening-in-de-wereld-van-eHealth.pdf [accessed 2019-10-11]
36. 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. EFI 2018 Dec 18;34(4):285-291. [doi: 10.3233/EFI-180221]
37. Higgins J, Green S. Cochrane handbook for systematic reviews of interventions. Vol 4: John Wiley & Sons; Feb 1, 2011.
38. Thomas B, Ciliska D, Dobbins M, Micucci S. A Process for Systematically Reviewing the Literature: Providing the Research Evidence for Public Health Nursing Interventions. Worldviews on Evidence-Based Nursing 2004 Jun 23;1(3):176-184. [doi: 10.1111/j.1524-475x.2004.04006.x]
39. 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(1):8 [FREE Full text] [doi: 10.1186/1471-2288-8-8] [Medline: 18298827]
40. Peel NM, Paul SK, Cameron ID, Crotty M, Kurrle SE, Gray LC. Promoting Activity in Geriatric Rehabilitation: A Randomized Controlled Trial of Accelerometry. PLoS One 2016 Aug 26;11(8):e0160906 [FREE Full text] [doi: 10.1371/journal.pone.0160906] [Medline: 27564857]
41. Vanoglio F, Bernocchi P, Mulè C, Garofali F, Mora C, Taveggia G, et al. Feasibility and efficacy of a robotic device for hand rehabilitation in hemiplegic stroke patients: a randomized pilot controlled study. Clin Rehabil 2017 Mar 10;31(3):351-360. [doi: 10.1177/0269215516642606] [Medline: 27056250]
42. Hesse S, Heß A, Werner C, Kabbert N, Buschfort R. Effect on arm function and cost of robot-assisted group therapy in subacute patients with stroke and a moderately to severely affected arm: a randomized controlled trial. Clin Rehabil 2014 Jul 22;28(7):637-647. [doi: 10.1177/0269215513516967] [Medline: 24452706]
43. Taveggia G, Borboni A, Mulé C, Villafañe JH, Negrini S. Conflicting results of robot-assisted versus usual gait training during postacute rehabilitation of stroke patients: a randomized clinical trial. Int J Rehabil Res 2016 Mar;39(1):29-35 [FREE Full text] [doi: 10.1097/MRR.0000000000000137] [Medline: 26512928]
44. Cannell J, Jovic E, Rathjen A, Lane K, Tyson AM, Callisaya ML, et al. The efficacy of interactive, motion capture-based rehabilitation on functional outcomes in an inpatient stroke population: a randomized controlled trial. Clin Rehabil 2018 Feb 19;32(2):191-200 [FREE Full text] [doi: 10.1177/0269215517720790] [Medline: 28719977]
45. Schoone M, Dusseldorp E, van den Akker-van Marle M, Doornebosch A, Bal R, Meems A, et al. Stroke Rehabilitation in Frail Elderly with the Robotic Training Device ACRE: A Randomized Controlled Trial and Cost-Effectiveness Study. Journal of Robotics 2011;2011:1-10. [doi: 10.1155/2011/543060]
46. Iosa M, Morone G, Fusco A, Castagnoli M, Fusco FR, Pratesi L, et al. Leap motion controlled videogame-based therapy for rehabilitation of elderly patients with subacute stroke: a feasibility pilot study. Topics in Stroke Rehabilitation 2015 Feb 25;22(4):306-316. [doi: 10.1179/1074935714z.0000000036]
47. Piqueras M, Marco E, Coll M, Escalada F, Ballester A, Cinca C, et al. Effectiveness of an interactive virtual telerehabilitation system in patients after total knee arthoplasty: a randomized controlled trial. J Rehabil Med 2013 Apr;45(4):392-396 [FREE Full text] [doi: 10.2340/16501977-1119] [Medline: 23474735]
48. Pol M, Ter Riet G, van Hartingsveldt M, Kröse B, Buurman B. Effectiveness of sensor monitoring in a rehabilitation programme for older patients after hip fracture: a three-arm stepped wedge randomised trial. Age Ageing 2019 Sep 01;48(5):650-657. [doi: 10.1093/ageing/afz074] [Medline: 31204776]
49. Oesch P, Kool J, Fernandez-Luque L, Brox E, Evertsen G, Civit A, et al. Exergames versus self-regulated exercises with instruction leaflets to improve adherence during geriatric rehabilitation: a randomized controlled trial. BMC Geriatr 2017 Mar 23;17(1):77 [FREE Full text] [doi: 10.1186/s12877-017-0467-7] [Medline: 28330455]
50. Laver K, George S, Ratcliffe J, Quinn S, Whitehead C, Davies O, et al. Use of an interactive video gaming program compared with conventional physiotherapy for hospitalised older adults: a feasibility trial. Disabil Rehabil 2012 Mar 12;34(21):1802-1808. [doi: 10.3109/09638288.2012.662570] [Medline: 22409245]
51. van den Berg M, Sherrington C, Killington M, Smith S, Bongers B, Hassett L, et al. Video and computer-based interactive exercises are safe and improve task-specific balance in geriatric and neurological rehabilitation: a randomised trial. J Physiother 2016 Jan;62(1):20-28 [FREE Full text] [doi: 10.1016/j.jphys.2015.11.005] [Medline: 26701163]
52. Bernocchi P, Giordano A, Pintavalle G, Galli T, Ballini Spoglia E, Baratti D, et al. Feasibility and Clinical Efficacy of a Multidisciplinary Home-Telehealth Program to Prevent Falls in Older Adults: A Randomized Controlled Trial. J Am Med Dir Assoc 2019 Mar;20(3):340-346. [doi: 10.1016/j.jamda.2018.09.003] [Medline: 30366759]
53. Barnason S, Zimmerman L, Schulz P, Tu C. Influence of an early recovery telehealth intervention on physical activity and functioning after coronary artery bypass surgery among older adults with high disease burden. Heart Lung 2009 Nov;38(6):459-468 [FREE Full text] [doi: 10.1016/j.hrtlng.2009.01.010] [Medline: 19944870]
54. Li CT, Hung GK, Fong KN, Gonzalez PC, Wah S, Tsang HW. Effects of a home-based occupational therapy telerehabilitation via smartphone for outpatients after hip fracture surgery: A feasibility randomised controlled study. J Telemed Telecare 2020 Jun 28:1357633X2093243. [doi: 10.1177/1357633x20932434]
55. Karner S, Stenner H, Spate M, Behrens J, Krakow K. Effects of a robot intervention on visuospatial hemineglect in postacute stroke patients: a randomized controlled trial. Clin Rehabil 2019 Dec 13;33(12):1940-1948. [doi: 10.1177/0269215519865993] [Medline: 31409126]

56. Franceschini M, Mazzoleni S, Goffredo M, Pournajaf S, Galafate D, Criscuolo S, et al. Upper limb robot-assisted rehabilitation versus physical therapy on subacute stroke patients: A follow-up study. J Bodyw Mov Ther 2020 Jan;24(1):194-198. [doi: 10.1016/j.jbmt.2019.03.016] [Medline: 31987544]
57. Da-Silva RH, Moore SA, Rodgers H, Shaw L, Sutcliffe L, van Wijck F, et al. Wristband Accelerometers to motiVate arm Exercises after Stroke (WAVES): a pilot randomized controlled trial. Clin Rehabil 2019 Aug 07;33(8):1391-1403. [doi: 10.1177/0269215519834720] [Medline: 30845829]
58. Ling Y, Ter Meer LP, Yumak Z, Veltkamp RC. Usability Test of Exercise Games Designed for Rehabilitation of Elderly Patients After Hip Replacement Surgery: Pilot Study. JMIR Serious Games 2017 Oct 12;5(4):e19 [FREE Full text] [doi: 10.2196/games.7969] [Medline: 29025696]
59. Backman C, Harley A, Kuziemsky C, Mercer J, Peyton L. MyPath to Home Web-Based Application for the Geriatric Rehabilitation Program at Bruyère Continuing Care: User-Centered Design and Feasibility Testing Study. JMIR Form Res 2020 Sep 14;4(9):e18169 [FREE Full text] [doi: 10.2196/18169] [Medline: 32924953]
60. White J, Janssen H, Jordan L, Pollack M. Tablet technology during stroke recovery: a survivor's perspective. Disabil Rehabil 2015 Sep 12;37(13):1186-1192. [doi: 10.3109/09638288.2014.958620] [Medline: 25212736]
61. Dakin L, Peel N. Effect of accelerometry on the functional mobility of older rehabilitation inpatients as measured by functional independence measure--locomotion (FIM) gain: a retrospective matched cohort study. J Nutr Health Aging 2011 May 30;15(5):382-386. [doi: 10.1007/s12603-010-0129-3] [Medline: 21528165]
62. Chan T, Chan F, Shea Y, Lin O, Luk J, Chan F. Interactive virtual reality Wii in geriatric day hospital: a study to assess its feasibility, acceptability and efficacy. Geriatr Gerontol Int 2012 Oct;12(4):714-721. [doi: 10.1111/j.1447-0594.2012.00848.x] [Medline: 22469232]
63. Hicks SA, Cimarolli VR. The effects of telehealth use for post-acute rehabilitation patient outcomes. J Telemed Telecare 2016 Dec 26;24(3):179-184. [doi: 10.1177/1357633x16686771]
64. Levinger P, Zeina D, Teshome AK, Skinner E, Begg R, Abbott JH. A real time biofeedback using Kinect and Wii to improve gait for post-total knee replacement rehabilitation: a case study report. Disabil Rehabil Assist Technol 2016 Sep 04;11(3):251-262. [doi: 10.3109/17483107.2015.1080767] [Medline: 26336875]
65. Goto K, Morishita T, Kamada S, Saita K, Fukuda H, Shiota E, et al. Feasibility of rehabilitation using the single-joint hybrid assistive limb to facilitate early recovery following total knee arthroplasty: A pilot study. Assist Technol 2017 Sep 30;29(4):197-201. [doi: 10.1080/10400435.2016.1219883] [Medline: 27689789]
66. Yoshikawa K, Mutsuzaki H, Sano A, Koseki K, Fukaya T, Mizukami M, et al. Training with Hybrid Assistive Limb for walking function after total knee arthroplasty. J Orthop Surg Res 2018 Jul 03;13(1):163 [FREE Full text] [doi: 10.1186/s13018-018-0875-1] [Medline: 29970139]
67. Koneva E, Timashkova G, Shapovalenko T, Lyadov K. Functional Spatially-Oriented Rehabilitation Of Elderly Patients After Cerebral Stroke. Reseasrch Journal of Pharmaceutical, Biological, and Chemical Sciences 2018;9(5):2232-2238.
68. Edmans J, Gladman J, Hilton D, Walker M, Sunderland A, Cobb S, et al. Clinical evaluation of a non-immersive virtual environment in stroke rehabilitation. Clin Rehabil 2009 Feb 01;23(2):106-116. [doi: 10.1177/0269215508095875] [Medline: 19164398]
69. Bernocchi P, Vanoglio F, Baratti D, Morini R, Rocchi S, Luisa A, et al. Home-based telesurveillance and rehabilitation after stroke: a real-life study. Top Stroke Rehabil 2016 Apr 09;23(2):106-115. [doi: 10.1080/10749357.2015.1120453] [Medline: 27078116]
70. Doornebosch AJ, Cools HJ, Slee-Turkenburg ME, van Elk MG, Schoone-Harmsen M. Robot-mediated ACtive REhabilitation (ACRE2) for the hemiplegic upper limb after a stroke: A pilot study. TAD 2008 Jan 03;19(4):199-203. [doi: 10.3233/tad-2007-19405]
71. Gandolfi M, Geroin C, Tomelleri C, Maddalena I, Kirilova Dimitrova E, Picelli A, et al. Feasibility and safety of early lower limb robot-assisted training in sub-acute stroke patients: a pilot study. Eur J Phys Rehabil Med 2017 Dec;53(6):870-882. [doi: 10.23736/s1973-9087.17.04468-9]
72. Hesse S, Waldner A, Tomelleri C. Innovative gait robot for the repetitive practice of floor walking and stair climbing up and down in stroke patients. J Neuroeng Rehabil 2010 Jun 28;7(1):30 [FREE Full text] [doi: 10.1186/1743-0003-7-30] [Medline: 20584307]
73. Sampson M, Shau Y, King MJ. Bilateral upper limb trainer with virtual reality for post-stroke rehabilitation: case series report. Disabil Rehabil Assist Technol 2012 Mar 29;7(1):55-62. [doi: 10.3109/17483107.2011.562959] [Medline: 21446826]
74. Schwickert L, Klenk J, Stähler A, Becker C, Lindemann U. Robotic-assisted rehabilitation of proximal humerus fractures in virtual environments: a pilot study. Z Gerontol Geriatr 2011 Dec 14;44(6):387-392. [doi: 10.1007/s00391-011-0258-2] [Medline: 22159833]
75. Marschollek M, Becker M, Bauer JM, Bente P, Dasenbrock L, Elbers K, et al. Multimodal activity monitoring for home rehabilitation of geriatric fracture patients--feasibility and acceptance of sensor systems in the GAL-NATARS study. Inform Health Soc Care 2014 Aug 22;39(3-4):262-271. [doi: 10.3109/17538157.2014.931852] [Medline: 25148561]
76. Tousignant M, Boissy P, Corriveau H, Moffet H. In home telerehabilitation for older adults after discharge from an acute hospital or rehabilitation unit: A proof-of-concept study and costs estimation. Disabil Rehabil Assist Technol 2006 Sep 09;1(4):209-216. [doi: 10.1080/17483100600776965] [Medline: 19260168]

77. Cimarolli VR, Reinhardt JP, Minahan J, Burack O, Thomas C, Melly R. Use of an Exercise Technology in Post-Acute Care of a Skilled Nursing Facility: A Feasibility Study. J Am Med Dir Assoc 2017 Nov 01;18(11):991.e1-991.e4. [doi: 10.1016/j.jamda.2017.07.002] [Medline: 28843523]
78. Peel NM, Russell TG, Gray LC. Feasibility of using an in-home video conferencing system in geriatric rehabilitation. J Rehabil Med 2011 Mar;43(4):364-366 [FREE Full text] [doi: 10.2340/16501977-0675] [Medline: 21305228]
79. Takano E, Ozaki K, Satoh K, Kawamura K, Maltais M, Kondo I. Effects of a Balance Exercise Assist Robot on Older Patients with Hip Fracture: A Preliminary Study. J. Med. Biol. Eng 2020 Oct 08;40(6):783-789. [doi: 10.1007/s40846-020-00568-x]
80. Mehrholz J, Pohl M, Platz T, Kugler J, Elsner B. Electromechanical and robot-assisted arm training for improving activities of daily living, arm function, and arm muscle strength after stroke. Cochrane Database Syst Rev 2018 Sep 03;9:CD006876. [doi: 10.1002/14651858.CD006876.pub5] [Medline: 30175845]
81. Cooper C, Gross A, Brinkman C, Pope R, Allen K, Hastings S, et al. The impact of wearable motion sensing technology on physical activity in older adults. Exp Gerontol 2018 Oct 02;112:9-19 [FREE Full text] [doi: 10.1016/j.exger.2018.08.002] [Medline: 30103025]
82. Goode AP, Hall KS, Batch BC, Huffman KM, Hastings SN, Allen KD, et al. The Impact of Interventions that Integrate Accelerometers on Physical Activity and Weight Loss: A Systematic Review. Ann Behav Med 2017 Feb 26;51(1):79-93 [FREE Full text] [doi: 10.1007/s12160-016-9829-1] [Medline: 27565168]
83. Talboom-Kamp EPWA, Verdijk NA, Kasteleyn MJ, Harmans LM, Talboom IJSH, Numans ME, et al. Effect of a combined education and eHealth programme on the control of oral anticoagulation patients (PORTALS study): a parallel cohort design in Dutch primary care. BMJ Open 2017 Sep 27;7(9):e017909 [FREE Full text] [doi: 10.1136/bmjopen-2017-017909] [Medline: 28963313]
84. Wentzel J, van der Vaart R, Bohlmeijer ET, van Gemert-Pijnen JEWC. Mixing Online and Face-to-Face Therapy: How to Benefit From Blended Care in Mental Health Care. JMIR Ment Health 2016 Feb 09;3(1):e9 [FREE Full text] [doi: 10.2196/mental.4534] [Medline: 26860537]
85. van der Kleij RM, Kasteleyn MJ, Meijer E, Bonten TN, Houwink EJ, Teichert M, et al. SERIES: eHealth in primary care. Part 1: Concepts, conditions and challenges. Eur J Gen Pract 2019 Oct 10;25(4):179-189 [FREE Full text] [doi: 10.1080/13814788.2019.1658190] [Medline: 31597502]
86. Valenzuela T, Okubo Y, Woodbury A, Lord SR, Delbaere K. Adherence to Technology-Based Exercise Programs in Older Adults: A Systematic Review. J Geriatr Phys Ther 2018 Jun 29;41(1):49-61. [doi: 10.1519/JPT.0000000000000095] [Medline: 27362526]
87. Broekhuis M, van Velsen L, Hermens H. Assessing usability of eHealth technology: A comparison of usability benchmarking instruments. Int J Med Inform 2019 Aug;128:24-31 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.05.001] [Medline: 31160008]
88. 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]
89. Maramba I, Chatterjee A, Newman C. Methods of usability testing in the development of eHealth applications: A scoping review. Int J Med Inform 2019 Jun;126:95-104 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.03.018] [Medline: 31029270]

Abbreviations

ICF: International Classification of Functioning, Disability, and Health.
MMAT: Mixed Methods Appraisal Tool.
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-analyses.

Edited by G Eysenbach; submitted 31.08.20; peer-reviewed by N Skjaeret, M Broekhuis, S Bhattacharjya; comments to author 19.12.20; revised version received 11.02.21; accepted 16.05.21; published 19.08.21.

Please cite as:
Kraaijkamp JJM, van Dam van Isselt EF, Persoon A, Versluis A, Chavannes NH, Achterberg WP
eHealth in Geriatric Rehabilitation: Systematic Review of Effectiveness, Feasibility, and Usability
J Med Internet Res 2021;23(8):e24015
URL: https://www.jmir.org/2021/8/e24015
doi:10.2196/24015
PMID:34420918

©Jules J M Kraaijkamp, Eléonore F van Dam van Isselt, Anke Persoon, Anke Versluis, Niels H Chavannes, Wilco P Achterberg. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Review

Overview and Strategy Analysis of Technology-Based Nonpharmacological Interventions for In-Hospital Delirium Prevention and Reduction: Systematic Scoping Review

Chan Mi Kim[1], MSc; Esther M van der Heide[2], PhD; Thomas J L van Rompay[3], PhD; Gijsbertus J Verkerke[4,5], Prof Dr; Geke D S Ludden[1], Prof Dr

[1]Department of Design, Production, and Management, Faculty of Engineering Technology, University of Twente, Enschede, Netherlands
[2]Patient Care and Monitoring Department, Philips Research, Eindhoven, Netherlands
[3]Department of Communication Science, Faculty of Behavioral, Management, and Social Sciences, University of Twente, Enschede, Netherlands
[4]Department of Biomechanical Engineering, Faculty of Engineering Technology, University of Twente, Enschede, Netherlands
[5]Department of Rehabilitation Medicine, University of Groningen, University Medical Center Groningen, Groningen, Netherlands

Corresponding Author:
Chan Mi Kim, MSc
Department of Design, Production, and Management
Faculty of Engineering Technology
University of Twente
De Horst (Building 20)
Enschede
Netherlands
Phone: 31 53 489 8450
Email: c.m.kim@utwente.nl

Abstract

Background: Delirium prevention is crucial, especially in critically ill patients. Nonpharmacological multicomponent interventions for preventing delirium are increasingly recommended and technology-based interventions have been developed to support them. Despite the increasing number and diversity in technology-based interventions, there has been no systematic effort to create an overview of these interventions for in-hospital delirium prevention and reduction.

Objective: This systematic scoping review was carried out to answer the following questions: (1) what are the technologies currently used in nonpharmacological technology-based interventions for preventing and reducing delirium? and (2) what are the strategies underlying these currently used technologies?

Methods: A systematic search was conducted in Scopus and Embase between 2015 and 2020. A selection was made in line with the Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR). Studies were eligible if they contained any type of technology-based interventions and assessed delirium-/risk factor–related outcome measures in a hospital setting. Data extraction and quality assessment were performed using a predesigned data form.

Results: A total of 31 studies were included and analyzed focusing on the types of technology and the strategies used in the interventions. Our review revealed 8 different technology types and 14 strategies that were categorized into the following 7 pathways: (1) restore circadian rhythm, (2) activate the body, (3) activate the mind, (4) induce relaxation, (5) provide a sense of security, (6) provide a sense of control, and (7) provide a sense of being connected. For all technology types, significant positive effects were found on either or both direct and indirect delirium outcomes. Several similarities were found across effective interventions: using a multicomponent approach or including components comforting the psychological needs of patients (eg, familiarity, distraction, soothing elements).

Conclusions: Technology-based interventions have a high potential when multidimensional needs of patients (eg, physical, cognitive, emotional) are incorporated. The 7 pathways pinpoint starting points for building more effective technology-based interventions. Opportunities were discussed for transforming the intensive care unit into a healing environment as a powerful tool to prevent delirium.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42020175874; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=175874

(J Med Internet Res 2021;23(8):e26079) doi:10.2196/26079

KEYWORDS

intensive care unit; delirium; delirium prevention; delirium reduction; delirium treatment; technology; technology-based intervention; strategy; nonpharmacological; systematic scoping review

Introduction

Background

Delirium is an acute brain dysfunction with a disturbance in attention, awareness, and cognition [1], which is common, especially in critically ill patients. It occurs in about 40% of patients in the intensive care unit (ICU) [2], and in case of ventilated patients, the proportion goes up to about 65%-80% [2,3]. Nonetheless, delirium can be often underestimated due to differences in the severity of illness in populations and underrecognition of delirium [4,5]. Delirium is frequently associated with a significant increase in the ICU length of stay [6], risk of long-term cognitive impairments [7], and 6-month mortality rates after leaving the ICU [8]. At the organizational level, delirium is associated with an increase in the cost of ICU care [9], ICU readmission [10], and mental stress of ICU nurses who take care of patients with delirium [11]. At the societal level, delirium costs up to $152 billion per year in health care in the United States [12]. Prevention of delirium may be the most effective way to avoid these negative outcomes. In case of in-hospital patients, at least 30%-40% of delirium cases are preventable by reducing the risk factors of delirium [13,14]. There are 2 types of interventions used for preventing delirium: pharmacological and nonpharmacological interventions (eg, reorientation).

Nonpharmacological multicomponent interventions are recommended over pharmacological interventions as a safe and promising way of preventing delirium [15]. Nonpharmacological multicomponent interventions aim at reducing delirium risk factors that are modifiable by, for instance, promoting better sleep, early mobilization, and cognitive training/stimulation [15]. In order to support the implementation of nonpharmacological interventions, several methods [16,17] have been introduced. As an example, the ABCDEF bundle provides practical ways to provide optimal care of ICU patients [16]. Despite the benefits of using these methods [18], there are still barriers, as the way in which nonpharmacological interventions are applied is often highly dependent on the medical staff. Technology-based interventions might help to overcome these barriers. In this review, “technology” is defined as equipment that is used or developed based on scientific knowledge for practical purposes. Ironically, while the ICU room has been one of the most technology-intensive places in hospitals, when it comes to delirium prevention, the use of technology-based interventions in the ICU room often remains limited to, for example, the use of earplugs [19].

With recent developments in technology, there is large potential for the diversity of technology-based interventions. Examples include a modified ICU room design providing, among others, a personalized light therapy system and noise reduction to support cognitive stimulation and a normal sleep-wake cycle of patients [20]. Another example is an interactive app using a conversational agent, which provides patients social interactions and medical advice [21]. The app enables personalized care for patients without frequent physical visits of bedside nurses. These examples showcase some of the possibilities that recent developments in technology can bring. However, understanding how to optimize technology-based interventions is limited, and therefore, it is worthwhile to further and more systematically study the topic.

Prior Work

Recently, several (systematic) reviews have been performed to provide overviews and to highlight the potential of existing nonpharmacological interventions for preventing delirium [18,22-28]. However, none of the existing reviews focused exclusively on technology-based interventions for preventing delirium. Moreover, none of the previous reviews [25-27] provided information necessary for improving the design and development of new technology-based interventions. For instance, a vision outlining what is needed to build a better care environment for patients in the context of intensive care and clear guidance for technology-supported improvements are lacking.

Goal of This Study

To close this gap, this paper will provide a comprehensive review of technology-based interventions used for preventing and reducing in-hospital delirium. To our knowledge, this is the first systematic effort to provide such an overview. Taking a bottom-up approach, we will first investigate what types of technologies are used in current nonpharmacological interventions and how these interventions contribute to delirium prevention and reduction in hospitalized patients. Second, we will identify underlying strategies of technologies applied in the technology-based interventions. Finally, following our analysis, we will discuss the limitations of the current technology use and the opportunities for further research and development of technology-based interventions aimed at delirium prevention in the ICU and for other hospitalized patients. This systematic scoping review was conducted to answer the following questions: (1) what are the technologies currently used in nonpharmacological technology-based interventions for preventing and reducing in-hospital delirium? and (2) what are the strategies underlying these currently used technologies? This review will provide insights into technologies that can be used for in-hospital delirium prevention and reduction and suggest directions for future design and development of innovative technology-based interventions. We propose that incorporating these insights will optimize the use of technologies and enhance the effectiveness of technology-based interventions.

Methods

Search Strategy

This systematic scoping review was conducted complying with the Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) guidelines [29]. The study protocol was registered on the international prospective register of systematic reviews (PROSPERO, CRD42020175874). The search was conducted from the Scopus and Embase databases for studies that are written in English and were published between January 1, 2015 and January 6, 2020 (Scopus)/January 13, 2020 (Embase). This time period was set to focus on the state-of-the-art technologies used following our scope for this review. The initial search was conducted with the terms delirium (deliri*) and technology (technolog*) but it did not result in a sufficient number of relevant studies. Samples of search terms were identified through previous studies and extended through the search of index terms, medical subject headings, and other technological terms used for other medical purposes. The following terms were used for the final search: (1) deliri*, (2) technolog*, (3) intelligen* OR automat* OR digital* OR computer OR computing OR robot*, (4) mobile OR app, (5) visual OR virtual OR VR OR video, (6) light* OR ambien* OR aroma* OR architect*, (7) sound* OR music* OR voice OR alarm, (8) cognitive training OR tracking OR game*, and (9) 1 AND 2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 (see Multimedia Appendix 1 for the full search strategy). Additional studies were sought through communication with experts. Only studies that were peer reviewed and conducted with human subjects were included. The title-abstract screening and the full-text review were performed by 2 reviewers (CMK and EMvdH) independently. Disagreements were discussed and reevaluated based on the main goal of the study. The reasons for excluding were recorded.

Eligibility Criteria and Study Selection

The studies were included based on the eligibility criteria (See Table 1).

Table 1. Eligibility criteria.

Criterion	Inclusion criteria	Exclusion criteria
Study type	All types of original research published in peer-reviewed journals and conferences	Letters, comments, editorials, conference abstracts, or any type of review
Period	January 1, 2015 until January 6 (Scopus) and 13 (Embase) 2020	Before January 1, 2015 and after January 6 (Scopus), 2020 and January 13 (Embase), 2020
Language	English	All other languages
Population	In-hospital patients in all age groups	Patients who are not in-hospital patients, healthy participants
Intervention	Nonpharmacological interventions using any type of technology either as a single intervention or as a part of multicomponent intervention aiming at preventing or reducing delirium	Pharmacological interventions, interventions only focusing on detecting/screening delirium
Comparator	Any comparator, including no comparator	N/A[a]
Outcome	Delirium-related data (eg, occurrence, duration of delirium) or risk factor–related data that indirectly influence delirium (eg, anxiety)	Neither delirium-related data nor risk factor–related data

[a] N/A: not applicable.

The main goal of this review was to find papers dealing with the prevention and reduction of delirium in the ICU, but in order to not overlook the potential of a greater range of technology-based interventions, the search scope was not limited to the ICU department but included also other hospital departments (eg, pediatric ICU, geriatric ward). We also included delirium from all age groups: from acute pediatric delirium to geriatric delirium because they share a similar range of delirium symptoms [30]. Moreover, the risk factors and recommended interventions across delirium in these groups are more or less the same [31]. Although delirium consists of 3 subtypes, each of which have their own symptoms and courses [32], these subtypes were not applied in the eligibility criteria and the study selection, as most studies did not specify them. We included studies focusing on incident delirium as well as prevalent delirium in order to address the full scope of delirium interventions. As we intended to explore all existing technology-based interventions supported by scientific evidence, we did not place any inclusion restrictions on the study design.

Data Extraction and Quality Assessment

Data extracted from the studies included the primary author, year of publication, country of origin, publisher, summary of intervention content, applied technology, intervention goal, study design, type and number of participants, outcome measure, intervention outcomes, key findings, and limitations of the study. The primary data extraction was performed by CMK using the predesigned data extraction form, and the extracted data were reviewed and confirmed by EMvdH. Disagreements were resolved by discussions between CMK and EMvdH, which included revisits of the relevant data by both authors. To distinguish differences in the strength of evidence in the studies, a quality assessment was performed by CMK using a predesigned assessment form (Multimedia Appendix 2) and reviewed and confirmed by EMvdH. Disagreements were resolved by follow-up discussions. However, none of the studies were excluded based on the quality assessment in our analysis as our goal in this review was to create an overview of the existing technology-based interventions.

Data Analysis

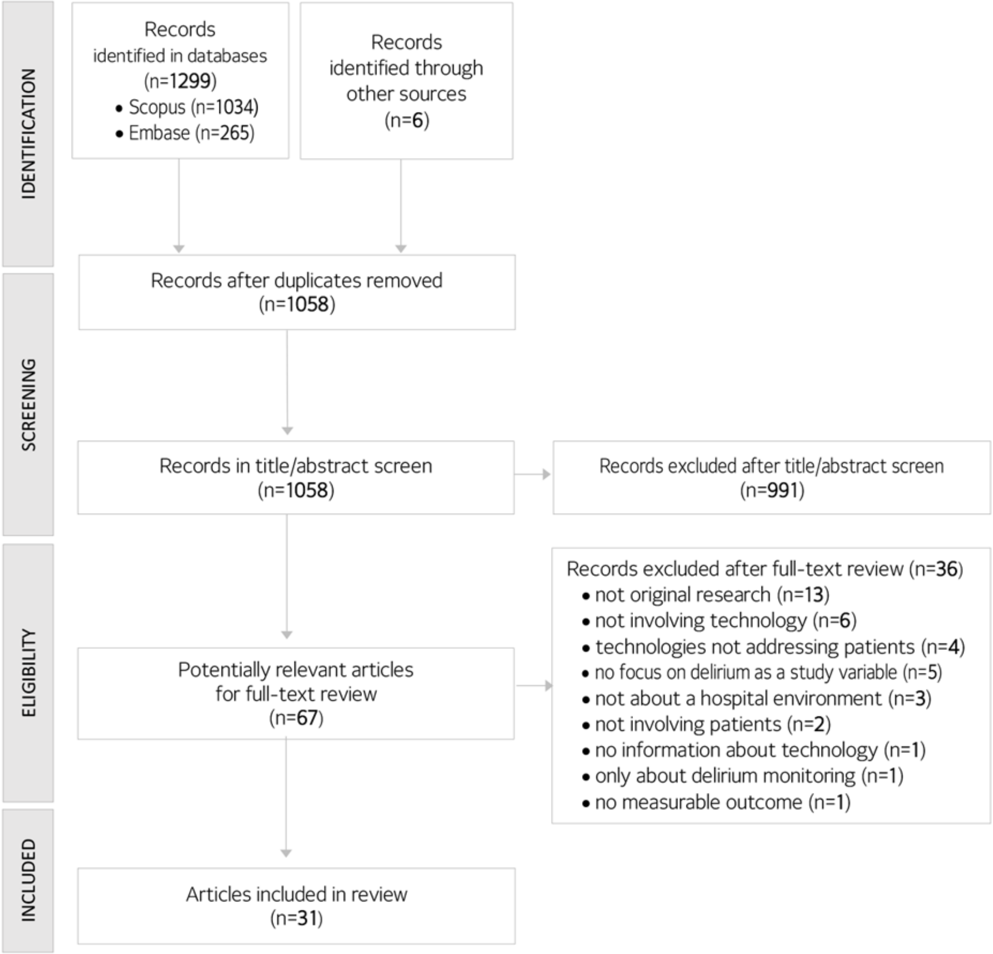
First, all technologies used in the technology-based interventions were looked into and clustered per technology type; a total of 8 categories was identified. Clustering was performed by CMK and reviewed and confirmed by EMvdH. Next, to summarize the strategies used in the technology-based interventions, goals and content summaries of each intervention were looked into, and 14 strategies were identified. These strategies were clustered and labelled based on the overarching theme: this led to the identification of 7 pathways to delirium prevention. An extraction of 14 elements from each intervention was carried out by CMK and clustering was performed independently by CMK and 2 other experienced researchers. After each session, the results and disagreements were discussed by CMK with the other experienced researchers, and then the synthesis was finalized by CMK.

Results

Study Selection

The process of literature screening and selection is presented in the flow diagram (Figure 1). After removing the duplicates, a total of 1058 studies were screened. Among these, 67 studies were examined based on the full text and 31 studies met inclusion for the review.

Figure 1. Flow diagram of study selection.


IDENTIFICATION
SCREENING
ELIGIBILITY
INCLUDED
Records identified in databases (n=1299)
• Scopus (n=1034)
• Embase (n=265)
Records identified through other sources (n=6)
Records after duplicates removed (n=1058)
Records in title/abstract screen (n=1058)
Records excluded after title/abstract screen (n=991)
Records excluded after full-text review (n=36)
• not original research (n=13)
• not involving technology (n=6)
• technologies not addressing patients (n=4)
• no focus on delirium as a study variable (n=5)
• not about a hospital environment (n=3)
• not involving patients (n=2)
• no information about technology (n=1)
• only about delirium monitoring (n=1)
• no measurable outcome (n=1)
Potentially relevant articles for full-text review (n=67)
Articles included in review (n=31)

Study Characteristics

The characteristics of the 31 included studies are listed in Table 2. Most of the studies were conducted in the ICU (15/31, 48%) followed by hospital wards (13/31, 42%) and both ICU and hospital wards (3/31, 10%). The most common study population was adult patients (23/31, 74%). Seven studies were conducted with pediatric patients [33-39] and 1 study was conducted with both adult and pediatric patients [40]. Most studies were prospective studies (28/31, 90%), including randomized trials [33-49], pretest-posttest/experimental [19,21,24,50-52], organizational case study [53], observational cohort [54], and pilot studies [55-57]. There were 3 retrospective studies [58-60] (See Table 2 for further specifications).

Table 2. Characteristics of the included studies classified into 8 different technology types.

Technology type, study	Main goal of intervention	Study design	Patient type[a]	Total number of patients (per group)
Audio: music/voice message				
Damshens et al [40]	To improve mental state[b]	Randomized clinical trial	ICU[c] (pediatric and adult)	80 (I[d]: 40, C[e]: 40)
Lee et al [39]	To reduce stress and anxiety	Randomized controlled trial	ICU	85 (I: 41, C: 44)
Johnson et al [48]	To alter physiologic response	Randomized controlled trial	ICU/trauma orthopedic unit	40 (I: 20, C: 20)
Sharda et al [53]	To mitigate postoperative pain and anxiety[b]	Organizational case study	Perioperative optimization of senior health	202 (I: 45, C: 157)
Cheong et al [55]	To enhance engagement and mood and to improve agitated behavior[b]	Pilot study (for randomized controlled trial)	Acute care unit (patients with delirium, dementia, or both)	25 (I: 25, C: 25)
Byun et al [33]	To activate positive psychological and behavioral responses and to reduce anxiety[b]	Double-blind randomized controlled trial	PACU[f] (pediatric)	66 (I1: 33, I2: 33)
Munro et al [49]	To support reorientation and to comfort patients[b]	Three-group prospective randomized controlled trial	ICU[g]	30 (I1:10, I2:10, C: 10)
Light: Dynamic light/natural light				
Estrup et al [58]	To improve circadian rhythm[b]	Retrospective cohort study	ICU	183 (I: 46, C: 137)
Pustjens et al [60]	To improve circadian rhythm[b]	Retrospective observational cohort study	Coronary care unit	748 (I: 369, C: 379)
Simons et al [41]	To improve circadian rhythm and sleep[b]	Randomized controlled single-center trial	ICU	734 (I: 361, C: 373)
Potharajaroen et al [42]	To improve sleep-wake cycle[b]	Single-blind randomized controlled study	ICU	62 (I: 31, C: 31)
Smonig et al [54]	To reduce circadian rhythm disruption[b]	Prospective single-center observational study	ICU	179 (I: 102, C: 77)
Video/video game: information/distraction				
Lee et al [50]	To reduce preoperative anxiety[b]	Quasi-experimental	ICU	50 (I: 25, C: 25)
Kim et al [34]	To reduce preoperative anxiety	Prospective randomized controlled trial	PACU (pediatric)	104 (I1: 34, I2: 33, I3: 37)
Rodriguez et al [35]	To reduce preoperative anxiety	Prospective randomized trial	PACU (pediatric)	52 (I1: 25, I2: 27)
Waszynski et al [43]	To reduce agitation in patients experiencing hyperactive or mixed delirium	Randomized controlled trial	Hospitalized/ICU[h]	111 (I1: 34, I2: 40, C: 37)
Dwairej et al [36]	To reduce preoperative anxiety	Randomized clinical trial	Day case surgery unit (pediatric)	128 (I: 64, C: 64)
Virtual reality: information/distraction				
Eijlers et al [37]	To reduce pain and anxiety[b]	Randomized controlled single-blind trial	Day case surgery unit (pediatric)	191 (I: 94, C: 97)
Ryu et al [38]	To reduce preoperative anxiety	Randomized controlled trial	PACU (pediatric)	86 (I: 41, C: 39)

Technology type, study	Main goal of intervention	Study design	Patient type[a]	Total number of patients (per group)
Suvajdzic et al [57]	To reduce clinical anxiety and depression and to support sleep and relaxation[a]	Pilot test	ICU	10
Sleep aids				
Demoule et al [44]	To improve sleep quality	Randomized controlled trial	ICU	45 (I: 23, C: 22)
Van de Pol et al [19]	To improve sleep quality	Interrupted time series design (preintervention and postintervention)	ICU	421 (I: 210, C: 211)
Communication aids				
Garry et al [56]	To improve patient's psychosocial status[b]	Pilot prospective trial	ICU/coronary care unit	12
Bott et al [21]	To provide companionship, social support, and health information to decrease loneliness and depression[b]	Case control quasi-experimental pre-post (for randomized controlled trial)	Medical and surgical unit	95 (I: 41, C: 54)
Others				
Lin et al [46]	To make the children adapt to the experience of visual disturbance	Prospective blinded randomized trial	PACU (pediatric)	179 (I: 89, C: 90)
Giraud et al [45]	To support mental status, earlier physical mobilization, and multisensory feedback/integration	Pilot time-cluster randomized controlled study	ICU	223 (I: 115, C: 108)
Multiple components				
Arbabi et al [51]	To support reorientation, cognitive/physical activation, human interaction[b]	Quasi-experimental	ICU	148 (I: 78, C: 69)
Tovar et al [52]	To reduce environmental precipitating factors, which impair sleep and to support maintenance of circadian cycle.	Prospective pre-experimental	ICU	49
Rivosecchi et al [24]	To provide cognitive stimuli and to support reorientation[b]	Prospective, observational quality improvement project pre-post	Medical ICU	483 (I: 253, C: 230)
Mitchell et al [47]	To support orientation and to provide cognitive stimulation	Single center randomized controlled trial	ICU	61 (I: 29, C: 32)
Zachary et al [59]	To prevent functional/cognitive, sleep and visual/hearing impairment, and dehydration	Retrospective	Medical, surgical/telemetry units[i]	4850 (I: 2146, C: 2704)

[a] All patients are adults unless mentioned as pediatric patients.

[b] With our minimal interpretation.

[c] ICU: intensive care unit.

[d] I: intervention group.

[e] C: control group.

[f] PACU: post anesthesia care unit.

[g] About 10% was delirious at admission.

[h] Only delirious patients.

[i] Excluding intensive care unit patients.

Quality Scores of the Included Studies

The assessed quality scores ranged from 1 [24,56] to 5 [32,36-39,46,48] (Multimedia Appendix 2). The average score was 3.258 (SD 1.264). A score lower than 3 was found in 10 studies [21,24,51-53,55-58,60]. Seven studies reached the maximum score of 5 [33,36-39,46,48]. The included studies were analyzed in 2 ways: first, by focusing on extracting the different types of technologies in the interventions, and second, by focusing on identifying strategies underlying these interventions.

Types of Technologies Currently Used for Preventing Delirium

A total of 31 technology-based interventions were identified (Table 2 and Table 3). Eight categories were distinguished based on the type of technology used. These categories are audio (7 studies), light (5 studies), video/video game (5 studies), virtual reality (VR) (3 studies), sleep aids (2 studies), communication support (2 studies), and others (2 studies). Technologies used as part of a multicomponent approach were grouped into a category called multiple components (5 studies). The identified technologies varied from simple (eg, earplugs, window blinds) to more advanced (eg, dynamic light, VR) ones. These technologies were used to complement conventional delirium treatments rather than replace them by reducing the negative (psychosocial) consequences of environmental factors and their effects on patient experiences.

Table 3. The content and effects of technology-based nonpharmacological interventions.

Study	Summary of the intervention	Incidence of delirium, *P* value	Duration of delirium (days)	Effectiveness		Key findings
				Direct[a]	Indirect[b]	
Audio: music/voice message						
Damshens et al [40]	*Music therapy*: Listening to light instrumental music selected by a music expert for 45 minutes twice (once in the morning and once at night) for a day.	I[c]: 15 (37.5%) C[d]: 16 (40%) (*P*=.82)	N/A[e]	No[f]	Yes[g]	No significant difference between the control and test group in terms of incidence of delirium. Significantly lower use of 2 pain relievers (acetaminophen and diclofenac) in the test group.
Lee et al [39]	*Music intervention*: A single 30-minute session of listening to slow-beat music (60-80 bpm) selected by patients from predefined playlists through headphones in the presence of a nurse at the bedside. The study was conducted to explore the anxiety-reducing effect of the intervention.	N/A	N/A	N/A	Yes	Music intervention significantly reduced anxiety and stress-related measures (serum cortisol level, heart rate, visual analog scale for anxiety, etc) of mechanically ventilated intensive care unit patients.
Johnson et al [48]	*Music intervention*: Listening to simple repetitive, self-selected music with slow tempo (60-80 bpm), low pitch, and repetitive rhythms for 60 minutes using an iPod and headsets, twice a day, over 3 days following admission.	I: 0 C: 0	N/A	N/A	Yes	No delirium incidence in both the control and test group. Music intervention significantly improved the pathophysiological mechanisms that contribute to delirium, neurotransmitter imbalance, inflammation, and acute physiological stressors in the test group.
Sharda et al [53]	*Confusion avoidance led by music program*: On postoperative day 1, patients got a music player and headphones with a personalized playlist based on a music assessment. Listening for at least 20 minutes twice a day was recommended, but patients had autonomy over the ultimate dose and frequency.	I: 17.8% C: 28.7% (*P*=.14)	N/A	No	No	This program may impact incident delirium and optimize postoperative pain and anxiety.
Cheong et al [55]	*Creative music therapy*: Spontaneous music making with musical instruments, playing familiar songs of patient's choice and music listening for 30 minutes once a day over 2 days. Based on individual’s profile and response to music, a certified music therapist modified techniques to meet patients’ need.	N/A	N/A	N/A	Yes	This therapy can improve mood/emotion and engagement of patients with dementia and delirium.
Byun et al [33]	*Mother's recorded voice*: Listening to either the recorded a voice of mother (I1) or a stranger (I2) through noise-cancelling headphones at the end of a surgery. The prerecorded message with standardized text was repeated with 10-second intervals and was continued until entering the postanesthesia care unit.	I1: 8 (24.2%) I2: 20 (60.6%) (*P*=.006)	N/A	Yes	N/A	Letting children listen to the sound of their mother in the recovery room reduced the incidence of emergence delirium.
Munro et al [49]	*Automated reorientation*: During the first 3 days at the intensive care unit, patients either received automated reorientation messages with a familiar (I1) or unfamiliar voice (I2) or no reorientation message at all (C). The message containing patient’s name and information about the intensive care unit environment (<2 minutes long) was played during the day (9 AM to 4 PM).	Delirium-free day I1: 1.9 days (0.99) I2: 1.6 days (1.07) C: 1.6 days (1.13) (*P*=.04)	I1:0.3 (0.48) I2:0.6 (0.84) C: 0.9 (1.28) (not significant)	Yes	N/A	Reorientation through automated messages increased the number of delirium-free days. A familiar voice was more effective than an unfamiliar voice in reducing delirium.

Study	Summary of the intervention	Incidence of delirium, *P* value	Duration of delirium (days)	Effectiveness		Key findings
				Direct[a]	Indirect[b]	
Light: dynamic light/natural light						
Estrup et al [58]	*Circadian light*: Exposure to the circadian light in which the amount of blue light (460-480 nm) changes over time like natural light. The light intensity varied from 50 lux to 4000 lux during daytime (between 6:00 and 20:30) and there was no blue light between 23:00 and 6:00. The exposure time of patients was varied from at least 24 hours to their total stay.	I: 30 C: 28	I: 2.5 (1-7) C: 1.5 (1-3) (*P*=.41)	No	N/A	Circadian light did not have an effect on the development of delirium.
Pustjens et al [60]	*Dynamic light*: Exposure to an artificial daylight system with light intensity peak values of 750 lux at the eye level and a color temperature ranging from 2700 K to 6550 K. Patients were exposed as long as possible during daytime. The overall period varied from 20 hours to 42.7 hours depending on the patient.	I: 20 (5.4%) C: 19 (5.0%) (*P*=.80)	N/A	No	N/A	Exposure to dynamic light did not reduce the incidence of delirium nor total hospital length of stay.
Simons et al [41]	*Dynamic light*: Exposure to the circadian light system with light intensity and color temperature peaks of 1700 lux and 4300 K via conventional fluorescent tubes between 9:00 and 16:00 except for 11:30 and 13:30 (intervention group) and to the standard lighting settings of 300 lux and 3000 K (control group) during the intensive care unit stay of patients (3-9 days).	I: 137 (38%) C: 123 (33%) (*P*=.16)	I: 2 (2-5) C: 2 (1-5) (*P*=.87)	No	N/A	Dynamic light as a single intervention did not reduce the cumulative incidence and duration of delirium in the test group.
Potharajaroen et al [42]	*Bright light*: Treatment with bright light therapy consisting of 5000 lux at 1.4-m distance from patient's face between 9:00 and 11:00 for 3 days. Other treatment data (nasal cannula oxygen, drugs etc) were analyzed.	I: 2 (6.45%) C: 11 (35.48%) (*P*=.005)	N/A	Yes	N/A	Bright light therapy reduced the incidence of delirium in the test group.
Smonig et al [54]	*Natural light exposure:* Exposure to natural light via windows from admission to the intensive care unit until discharge (3-7 days).	I: 65 (64%) C: 55 (71%) (*P*=.28)	I:3 (1-6) C:3 (1-7) (*P*=.43)	No	Yes	Admission to a single room with natural light via windows did not reduce the incidence of delirium but a risk of agitation episodes and hallucinations
Video/video game: information/distraction						
Lee et al [50]	*Preoperative video information*: Informative videos that explain preoperative procedures, operating room environment, and intensive care unit environment on the day prior to surgery.	I: 3 (12%) C: 5 (20%) (*P*=.26)	N/A	No	Yes	The intervention was not effective in reducing delirium incidence but decreased the anxiety levels.
Kim et al [34]	*Video distraction and parental presence*: In the preoperative holding room before surgery, provision of a 4-minute animated cartoon video (I1), parental presence (I2), or a video plus parental presence (I3). The primary study goal was to compare the effect of video distraction, parental presence, or combination of both on the preoperative anxiety reduction.	I1: 13 (38.2%) I2: 13 (39.4%) I3[e]: 20 (43.5%) (*P*=.32)	N/A	No	N/A	All groups showed similar effect on the preoperative anxiety and none of them significantly reduced emergence delirium.

Study	Summary of the intervention	Incidence of delirium, *P* value	Duration of delirium (days)	Effectiveness		Key findings
				Direct[a]	Indirect[b]	
Rodriguez et al [35]	*Video distraction using varying screen size*: Watching a movie in the preoperative area and through the induction of anesthesia. Patients chose from one of the 5 preselected age-appropriate movies using either a large bedside screen (I1) or a small tablet (I2). One parent accompanied a patient. The average time spent was 3.8 minutes (I1) and 4.5 minutes (I2). The primary study goal was to compare the effect of video distractions in different screen sizes on anxiety reduction.	I1: 29.16% I2: 30.8% (not significant)	N/A	No	Yes	The video distractions decreased the preoperative anxiety, regardless of the size, with parental presence at induction of anesthesia. No effect was found for the emergence delirium rate.
Waszynski et al [43]	*Simulated family presence and nature scene*: Two types of video intervention for when agitation is present and the family is not: watching a 1-minute family video message plus usual care (I1) or watching a 1-minute nature video plus usual care (I2). The study goal was to examine the effect of family video message on agitation level.	N/A	N/A	N/A	Yes	Both family video message and nature video can decrease agitation in delirious patients.
Dwairej et al [36]	*Video game distraction and anesthesia mask practice*: Combination of video distraction using a handheld video game (1-2 minutes) before the transfer to operation room, anesthesia mask exposure, and shaping intervention. During anesthesia induction, parental presence is allowed but not standardized. In the operating room, nonmedical talks occurred to distract the child. The study goal was to evaluate the effectiveness of the intervention on the preoperative anxiety.	I: mean 11.06, SD 3.97 C: mean 10.25, SD 4.81 (*P*=.30)	N/A	No	Yes	The intervention significantly reduced anxiety. Yet, the results did not reveal statistically significant difference in emergence delirium scores.
Virtual reality: information/distraction						
Eijlers et al [37]	*Virtual reality exposure*: Provision of a 15-minute highly immersive virtual reality experience of the operating theatre to get familiarized with the environment and general anesthesia procedures. The virtual environment was computer-generated, interactive, and child-friendly.	I: 7.0 (5.0-9.0)[f] C: 6.0 (5.0-9.0) (*P*=.27)	N/A	No	No	This did not have a beneficial effect on anxiety, pain, emergence delirium, or parental anxiety.
Ryu et al [38]	*Preoperative immersive virtual reality tour of operating theater*: Provision of a 4-minute virtual reality video for pediatric patients showing the operating theater and explaining the perioperative process by using a popular animal character as a patient. The intervention was provided 1 hour prior to entering the operating room. The study goal was to examine the effect of the intervention on reducing the preoperative anxiety.	I: 16 (39%) C: 14 (36%) (*P*=.77)	N/A	No	Yes	The intervention did not reduce the incidence and severity of emergence delirium, although it was effective in alleviating preoperative anxiety in children.

Study	Summary of the intervention	Incidence of delirium, *P* value	Duration of delirium (days)	Effectiveness		Key findings
				Direct[a]	Indirect[b]	
Suvajdzic et al [57]	*Patient-centered virtual reality system*: Intervention consisting of 2 sessions (session 2 was held at least 24 hours after session 1): (1) a video instructing patients to enjoy the movie by moving their heads to look around, followed by a 5-10-minute guided meditation in virtual nature scenes for breath control (Relax VR), (2) playing either Relax VR or fishing game (Bait!).	I: 0 (0%)	N/A	N/A	No	The interventions did not result in clinically significant changes in pain, sleep, or vital signs. It seems likely that greater exposure to virtual reality is more likely to produce a meaningful effect on patient physiology and sleep quality.
Sleep aids						
Demoule et al [44]	*Earplugs and eye masks*: Use of earplugs and eye masks every night between 10 PM and 8 AM from inclusion until ICU discharge (average 7 days). The study goal was to evaluate the impact of the intervention on sleep architecture in intensive care unit patients.	I: 2 (7%) C: 2 (6%) (*P*>.99)	N/A	N/A	Yes	Interventions resulted in reduced long awakenings and increased deep sleep duration. Possibly the effect was at least partially counteracted by the discomfort of wearing the devices.
Van de Pol et al [19]	*Nocturnal sound-reduction protocol*: A protocol focusing on reducing noise in the night, for example, speaking and laughing quietly in the lobby, minimizing alarm volume, closing the door when the patient is not delirious, and providing earplugs at night. One month of implementation phase.	Slope of delirium incidence I: –2.79% (*P*=.02) C: 0.91% (*P*=.37) Difference: –3.70% per time period (*P*=.02)	N/A	Yes	Yes	The protocol reduced the incidence of delirium. It significantly reduced delirium risk factors such as perceived nighttime noise and the use of sleep medication. Reported sleep quality was not improved.
Communication aids						
Garry et al [56]	*Eye-tracking devices*: Usage sessions (45 min/session, 5 sessions on consecutive weekdays) were given, during which patients were prompted to spell out notes, indicate their needs via picture sets, and play simple memory games. Patients were permitted to communicate with family, nursing staff, and physicians outside the training sessions.	Day 1: 4 (33%) Day 2: 1 (8%) Day 3/onward: 0	N/A	N/A	Yes	The use of an eye-tracking device positively affected patients' happiness and ability to participate; however, it did not show a significant effect on patients' confusion level or frustration.
Bott et al [21]	*Bedside digital care coach avatar*: 24-hour psychosocial and health care support through an embodied conversational agent with an appearance of animated animal avatar. It checks patient status, assists communication, and offers psychological support during their stay at medical and surgical units (3-6 days). The average time spent with the embodied conversational agent was 61 minutes per day.	I: pre12 (41%)/post1 (3%) C: pre6 (13%)/post3 (6%) (*P*<.01/.25)	N/A	Yes	Yes	The use of the care coach avatar during hospitalization can reduce the frequency of delirium, loneliness, and falls among diverse hospitalized older adults.
Others						
Lin et al [46]	*Eyepatch for visual preconditioning*: Preventive treatment for pediatric patients undergoing ophthalmic surgery consisted of covering the eye with an eyepatch for at least 3 hours one day before surgery.	I: 15 (16.9%) C: 40 (44.4%) (*P*<.001)	N/A	Yes	Yes	The intervention significantly reduced preoperative anxiety and emergence delirium. Preoperative anxiety was found to be an independent risk factor of emergence delirium.

Study	Summary of the intervention	Incidence of delirium, *P* value	Duration of delirium (days)	Effectiveness		Key findings
				Direct[a]	Indirect[b]	
Giraud et al [45]	*Structured mirrors intervention*: Protocol-driven mirrors intervention consisting of different mirrors to provide visual feedback about the environment as a re-orientation tool and to support self-awareness and explanation of medical/nursing procedures. The unit of randomization was a 2-week time period cluster.	I: 20 (17%) C: 17 (16%) (*P*=.71)	I: 1 (IQR 1-3 [range 1-25]) C: 2 (IQR 1-8 [range 1-13]) (*P*=.40)	No	Yes	Use of the mirror intervention did not reduce delirium but improved factual memory encoding.
Multiple components						
Arbabi et al [51]	*Environmental changes and liaison education*: Environment with proper time cues, appropriate lighting for the time of the day during intensive unit care stay (average about 5 days). Further, allowing interactions with family members and medical staff, giving vision and hearing aids, preventing dehydration, and encouraging early mobilization. Training for medical staff on delirium management.	I: 30 (37.97%) C: 50 (72.46%) (*P*=.01)	I: 26.18 (SD 35.38) C: 35.84 (SD 39.31) (*P*=.001)	Yes	N/A	Multifactorial intervention (educational and environmental changes) was effective in reducing the delirium rate in the intensive care unit and the duration of delirium.
Tovar et al [52]	*Environment with reduced environmental stressors*: Nursing care guide to reduce environmental stressors such as noise and continuous artificial light. Provision of active interactions with family members and medical staff, cognitive/sensory stimuli, and information about the environment. The posttest was performed after the guideline had been applied for 5 days.	I: 3 (6.12%)	N/A	Yes	N/A	Multicomponent intervention focusing on reducing precipitating environmental factors was effective in reducing delirium incidence and improving sleep.
Rivosecchi et al [24]	*Nonpharmacological protocol*: Nursing education bundled into the protocol consisting of music, opening blinds, re-orientation and cognitive stimulation, eye and ear protocol during intensive care unit stay (median 188.3 hours-control group, and 153.5 hours-intervention group).	I: 24 (9.4%) C: 36 (15.7%) (*P*=.04)	I:16 h (IQR 8-24) C:20 h (IQR 9.5-37) (*P*<.001)	Yes	N/A	The protocol reduced both duration and incidence of delirium.
Mitchell et al [47]	*Multicomponent family-delivered intervention*: Family intervention consisting of orientation (memory clues: family pictures etc), therapeutic engagement (cognitive stimulation: discussing current family life events etc), and sensory (glasses, hearing aids in place/working). The intervention group was enrolled for a median of 5 days and the family members were asked to deliver the intervention at least once a day. The study goal was to assess the feasibility and acceptability of the intervention for designing a larger randomized controlled trial.	I: 17 (59%) C: 18 (56%) (*P*=.87)	I: 1.0 (2) C: 1.0 (2) (*P*=.60)	No	N/A	Multicomponent family-delivered intervention was not effective in reducing delirium incidence nor days of delirium.

Study	Summary of the intervention	Incidence of delirium, *P* value	Duration of delirium (days)	Effectiveness		Key findings
				Direct[a]	Indirect[b]	
Zachary et al [59]	*Hospital elder life program interventions*: Geriatric intervention program consisting of reorientation, social stimulation, music therapy, games, mindfulness relaxation, mobilization, visual/hearing aids, sleep aids, and nutrition support. Volunteer interventions were 3 times a day, which took 20-30 minutes each time.	N/A	N/A	N/A	Yes	This program reduced 30-day readmissions and hospital length of stay in the 70-85 years age group.

[a] Direct effects: incidence and duration of delirium.

[b] Indirect effects: length of stay or delirium risk factors, including but not limited to pain, anxiety, stress, mood, agitation, and sleep deprivation.

[c] I: intervention group.

[d] C: control group.

[e] N/A: not applicable.

[f] No significant positive effect.

[g] Significant positive effect.

Of the 31 identified studies, 23 studies analyzed the effect of the interventions on direct delirium-related outcome measures (incidence and duration of delirium). The other 8 studies only assessed indirect outcome measures (eg, length of stay, precipitating risk factors of delirium such as anxiety). Of the 23 studies that evaluated direct delirium-related outcome measures, 9 studies showed a significant effect on decreasing either or both incidence and frequency of delirium. The interventions in these studies involved diverse technologies across the categories, that is, audio [33,49], light [42], sleep aids [19], communication support [21], others [46], and multiple components [24,51,52]. Of the 14 studies without significant effect on direct outcome measures, 8 studies, however, showed significant effects on delirium risk factors or symptoms, including pain [40], anxiety [35,36,38,50], agitation [54], hallucination [54], and factual memory encoding [45]. Six of the 8 studies using diverse technology-based interventions showed a significant effect on indirect outcome measures, including anxiety [39], agitation [43], sleep deprivation [44], or ICU readmission [59]. The technologies used in these studies were audio [39,48,55], video/video game [43], sleep aids [44], and multiple components [59]. It is notable that across all technology categories, significant positive effects were found on either or both direct and indirect delirium outcome measures. An explanation and further analysis of each category is given as follows.

Audio

The interventions using audio had either music or prerecorded voice messages. Music interventions involved, in general, slow tempo music with a duration of 20 minutes to 1 hour. The music was provided 2 or 3 times a day by using a television audio system or headset. These interventions aimed at addressing anxiety [39,53], mood [55], pain [40,53], and engagement [55] for adult patients. Across these interventions, we found differences in the level of personalization and autonomy: the level of personalization varied from the provision of music preselected by a music expert to the creation of a music playlist based on a patient's own choice and the level of autonomy ranged from patients listening to music with dose and frequency predecided by a researcher to patients deciding on dose and frequency. Interestingly, delirium incidence was reduced in the study with personalized music allowing patients autonomy [53]. While the majority of studies used a passive format such as music listening, 1 study used a participatory format providing an interactive music therapy and showed significant improvement in mood and engagement [55]. Prerecorded voice messages contained mainly patient's name and information about the care environment using either a familiar or unfamiliar voice. These interventions aimed to provide reorientation and a feeling of comfort. Two studies were carried out with different patient age groups: one with pediatric patients and the other with adult patients. The use of a prerecorded voice message reduced direct delirium outcome in both groups [33,49]. Likewise, one of the studies showed that a familiar voice was more effective than an unfamiliar one [49].

Light

The interventions involving light used either a specially made light therapy system for dynamic/bright light or natural light through windows. Dynamic light interventions provided light consisting of diverse intensities (ranging from 50 lux to 4000 lux) and color temperatures (ranging from 2700 K to 6550 K) in the environment of patients [41,58,60]. The bright light intervention used high intensity of light (5000 lux) for 2 hours a day [42] while the natural light intervention used natural light coming through windows [54]. All studies aimed at improving the patients' circadian rhythm and were only tested on adult patients. Only the bright light intervention showed an effect on reducing the incidence of delirium [42]. Possibly, most of the current light interventions are not effective in reducing direct delirium outcomes. Another possible explanation for why few studies found an effect on reducing incidence of delirium might relate to the amount of light (intervention) that patients actually received. In most studies, for example, the actual light intensity at patients' eye level was not specified or the definite exposure time to light of each patient was not guaranteed as sedative patients were included who had their eyes closed [54,58,60].

Video/Video game

The interventions using video/video games were used for providing either information about the medical procedure prior to a surgery or distractions with various contents, including age-appropriate programs (eg, cartoons), family messages, or nature scenes through a handheld tablet or a specially made bedside screen. These interventions aimed to address anxiety or agitation and were effective in decreasing anxiety or agitation in all studies with either pediatric or adult patients [34-36,43,50]. Regarding direct delirium outcome measures, none of the studies showed a significant effect [34-36,50].

VR Technology

The interventions using VR technology provided either information about medical procedures, distraction, or sensory stimulation for patients. VR interventions used a head-mounted device and the VR content varied from a guided tour to the operating theatre, to the virtual scenes of real-world locations. The interventions were mainly used to address anxiety or to support restorative effect for both pediatric and adult patients [37,38,57]. The effects of using VR to provide information (a preoperative VR tour to the operating theatre) differed in the studies. Of the two, 1 study showed a significant effect on reducing the anxiety of pediatric patients [38]. For none of these studies, the use of VR resulted in any effect on delirium-related outcome measures [37,38,57].

Sleep Aids

The interventions related to sleep aids were wearable devices such as earplugs and eye masks [19,44], and environmental modifications such as closing doors and window blinds were applied [19]. They aimed at reducing noise and light during nighttime to improve the sleep quality of adult patients. The application of both wearables and environmental modifications showed an effect on reducing the incidence of delirium [19]. Moreover, the use of a wearable showed an improvement in sleep quality [44]. It is notable that this study also highlights the potential negative side-effects of using wearables due to discomfort experienced by vulnerable patients.

Communication Supports

The technologies used for communication supports were a conversational agent [21] and an eye-tracking device [56]. They assisted adult patients to express their needs or to participate in psychosocial activities. Both types improved the psychological well-being of patients by increasing happiness and the ability to participate and by reducing loneliness. The conversational agent was used as a digital care coach providing communication means, human interactions, and companionship. This intervention reduced the frequency of delirium [21].

Others

In this category, we found rather simple forms of technologies used for diverse purposes: an eyepatch for experiencing what will happen after surgery [46] and a structured mirror to provide patients cognitive stimuli as a means to support mobilization and communication [45]. The eyepatch was used for pediatric patients and was effective in reducing emergence delirium [46]. The structured mirror intervention was for adult patients and improved their factual memory encoding [45].

Multiple Components

Some interventions involved more than one technology as part of a nonpharmacological bundle for adult patients. Environmental modifications for noise reduction, cognitive stimulation, and reorientation reduced delirium incidences in 3 studies [24,51,52]. Zachary et al [59] had a similar focus and showed an effect on reducing the 30-day readmission. In a small feasibility study, no significant effect was found on direct delirium outcome measures for the simple technology-based intervention involving families throughout different therapies such as orientation and cognitive stimulation [47].

Seven Pathways to Delirium Prevention

Fourteen strategies to prevent delirium were identified from the technology-based interventions of the included studies (see Multimedia Appendix 3 for the 14 strategies used in the included studies and see Table 4 for full description of the 14 strategies).

Table 4. Descriptions of the 14 strategies.

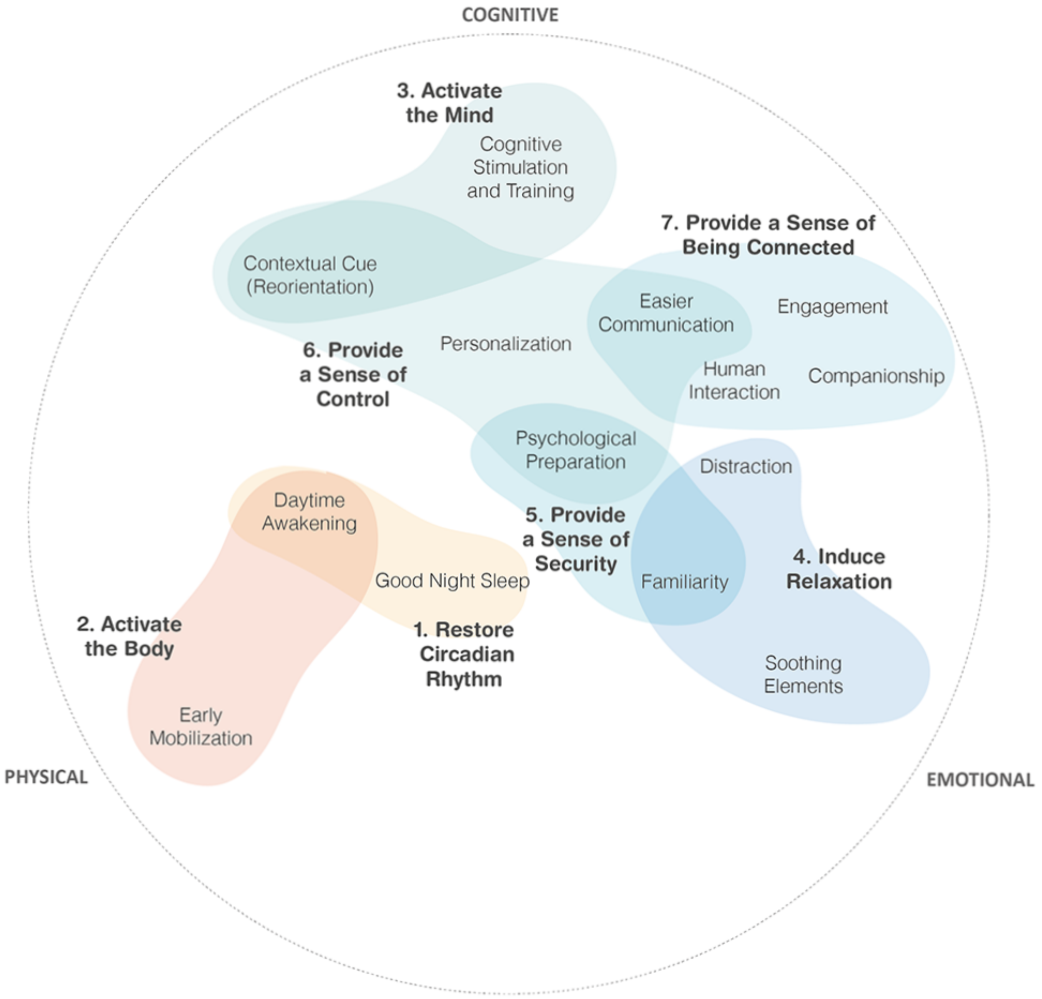
Strategy	Explanation	Example
Cognitive stimulation and training	Stimulating patient’s brain activity to maintain and improve their cognitive capability and executive functions such as attention, memory, reasoning, and language.	Music, book
Companionship	Providing patients a sense of consistent social presence as a means to combat social isolation and loneliness.	Digital agent
Contextual cue (reorientation)	Providing patients contextual information such as time, date, and place to minimize confusion and anxiety coming from not knowing what’s going on and feeling lost	Automated voice message, clock
Daytime awakening	Supporting patients to stay physically and mentally activated during day so they can become tired enough to sleep at night	Dynamic light
Distraction	Redirecting patient’s focus away from distressing situations/conditions such as pain, discomfort, fear, and anxiety	Music, video, virtual reality
Early mobilization	Encouraging patients to move their bodies early enough to prevent muscle loss and other complications caused by lack of physical movement	Structured mirror
Easier communication	Providing a means for patients to better express their needs especially when they are mechanically ventilated	Eye-tracking device
Engagement	Encouraging patients to be interested in and to be involved with what is happening	Participatory music therapy
Familiarity	Providing something that patients feel familiar with to help them feel safe, at ease, and calm	Mother’s voice, personalized music list
Good night sleep	Providing an environment that facilitates sleeping by removing disturbing elements such as sound and light noise and by adding elements enhancing patient comfort and relaxation	Ear plugs
Human (social) interaction	Providing patients warm human (-like) interactions to stimulate them socially and to help them feel being involved and being cared for	Digital agent
Personalization	Providing an option that reflects patient’s preference	Personalized music list
Psychological preparation	Helping patients to feel prepared and confident by informing them what will happen in advance.	Virtual tour to the operation room
Soothing elements	Helping patients to calm down and to manage stress and anxiety by providing an activity or environment that is soothing	Music, nature video

Subsequently, the 14 strategies were clustered into 7 pathways by using a thematic analysis approach (Figure 2). As such, the 7 pathways might provide directions toward technology-based interventions for delirium prevention. The 7 pathways and a short description of each pathway are as follows:

1. Restore the circadian rhythm: helping patients to find a normal sleep-wake cycle to prevent sleep deprivation.
2. Activate the body: supporting patients to regain physical strength and endurance.
3. Activate the mind: supporting patients to prevent cognitive decline, restore cognitive function, and minimize confusion, which can cause negative emotions such as anxiety, agitation, and aggression.
4. Induce relaxation: helping patients to stay in a positive psychological state, which prevents emotional distress, makes it easier to cope with their situation, and improves the patients’ physical state (eg, through better sleep).
5. Provide a sense of security: supporting patients in feeling reassured and safe so that they can easily handle stress and emotional distress such as anxiety and fear that originate mainly from uncertainty and unfamiliarity.
6. Provide a sense of control: supporting patients by enhancing autonomy, empowerment, and control over anxiety.
7. Provide a sense of being connected: supporting patients to feel connected and socially engaged to prevent loneliness, depression, and anxiety.

Three of the 7 pathways outlined above, that is, *Restore the Circadian Rhythm*, *Activate the Body*, and *Activate the Mind*, are in line with strategies recommended in the ABCDEF bundle [16]. The other pathways are not directly linked, yet are associated with important predictors of well-being used in psychology: *Induce Relaxation* links to coping strategies [61] and *Provide a Sense of Security*, *Control*, and *Being Connected* are related to the universal psychological needs, which are security [62], dominance [63], and relatedness [62]. The 7 pathways, therefore, cover a broad range of delirium prevention strategies correlating physical, cognitive, and emotional aspects as shown in Figure 2.

Figure 2. Overview of the 14 strategies grouped into 7 pathways of delirium prevention that are used in technology-based interventions.


COGNITIVE
3. Activate the Mind
Cognitive Stimulation and Training
Contextual Cue (Reorientation)
7. Provide a Sense of Being Connected
Easier Communication
Engagement
Personalization
6. Provide a Sense of Control
Human Interaction
Companionship
Psychological Preparation
Distraction
Daytime Awakening
5. Provide a Sense of Security
Good Night Sleep
Familiarity
4. Induce Relaxation
2. Activate the Body
1. Restore Circadian Rhythm
Soothing Elements
Early Mobilization
PHYSICAL
EMOTIONAL

Discussion

Overview

The results of this review provided an overview and characteristics of technology-based interventions that have been used to prevent and reduce delirium. We also summarized the related strategies into 7 pathways. These 7 pathways include key elements for developing technology-based interventions for delirium prevention. From our analysis of the included studies and technologies, we next discuss the limitations and opportunities for future research.

Limitations in the Current Technology Use for Preventing Delirium

First, most technology-based interventions addressed only one or a few delirium risk factors. For instance, interventions that used a dynamic light system [41,60] aimed to improve the patients' sleep-wake cycle while interventions, including video distraction [34,35,43], aimed to reduce the anxiety of the patients. Such approaches are not in line with the frequently recommended multicomponent approach, which effectively targets the multifactorial origin of delirium [15]. Second, most technology-based interventions were more momentary than continuous solutions. Although the length of stay in the ICU should be kept as short as possible, it can last from days to weeks. During this period, depending on the severity of their illness, patients are going through various clinical and emotional phases and their needs are changing respectively. During various activities (eg, clinical checkup, therapy session, cleaning, resting, family visit), patients can experience diverse feelings (eg, anxiety, fear, tiredness, worrying, relaxed, cheerful) and have different corresponding needs. In contrast to the heterogeneity of activities and patients' needs throughout their ICU stay, the majority of the included studies focused on a short period of time. For instance, the video and video game interventions were usually planned for less than 5 minutes [34-36,43,50], and the VR interventions were used for less than 15 minutes [37,38,57]. In order to minimize delirium risk factors, it is important to understand the needs and concerns of patients throughout their whole ICU stay and to develop preventive solutions that are more continuous and seamless. Importantly, knowledge as how to develop and implement a continuous solution throughout the ICU stay of patients is missing as none of the reviewed studies evaluated it. Third, only a limited number of studies focused on the use of technology for improving the patients' environment. Despite the negative influences of the stressors in the current ICU environment, such as overload of light and noise [64,65] on patients' clinical progress, only a few studies [41,51,52,58,60] aim at optimizing the ICU environment. In contrast, most technologies used as interventions added extra sensory burden to the patients. Last, the effect of some technologies can be improved by a more thorough understanding of (vulnerable) patients and context-dependent needs. Understanding the patient's needs is crucial for delivering high-quality care. However, sometimes, the implementation of technology can be challenging and might result in a suboptimal match with the patients and context. For instance, in the study of bright light therapy, patients were sedated and had their eyes closed and this resulted in insufficient amount of light exposure [41]. The use of wearable sleep-aid devices (earplugs and eye masks) by ICU patients can lead to potential discomfort, thereby limiting the effectiveness of the intervention [44].

Further Development of Technology-Based Interventions for Preventing Delirium

Despite the limitations of current technology-based interventions, the potential of technology in delirium prevention is promising [20]. Reflecting on the identified limitations, the following recommendations are proposed for the future design and development of technology-based interventions for delirium prevention. First of all, technology should not only support physical and cognitive functions but also support psychological and emotional needs. In the analysis, the proposed 7 pathways described existing approaches for delirium prevention, which cover multifaceted needs of patients. Yet, comparing the number of related strategies and studies included in each pathway, it is notable that, in general, there are far more technology-based interventions aiming to support functions than to support needs. To take a more comprehensive approach, further development of technologies should aim at meeting patients' needs by, for instance, providing a means to allow them more control over their situation, feel relaxed, safe, and connected. Some examples are an interaction device to enable ventilated patients to easily express their needs using gestures [66], a robotic pet that lets patients cuddle and helps them feel calm [67], or an intelligent alarm system contributing to a relaxing environment for ICU patients by controlling and harmonizing alarm sounds [68]. Another way to meet patients' emotional and psychological needs can be found in the provision of (multi-) sensorial and cognitive stimuli. For instance, aromatherapy can be a way to support patients to feel relaxed through sensorial stimuli combining pleasant tactile pressure and aromatic fragrance [69]. For the stimuli, using nature elements can be an interesting candidate as exemplified by previous cases [43,57] and other applications: examples are a VR therapy showing various nature sceneries [70] and a geriatric care environment adapting nature elements [71] to generate relaxation [70,71], increase social engagement, and reduce restlessness [71]. Second, technology could create a healing environment for patients: a more context-aware, personalized, and adaptive ICU. Despite emerging interest in patient-centered care, this review showed that such interactive technology is rare in the current ICU environment. Patient-centered care recognizes a patient as a unique individual and stresses the importance of care tailored to patients' specific preferences, needs, and values [72-74]. In order to better adapt patient-centered care, technology should be evolved in a way that (1) it allows patients to take a more active role in their care process, for instance, by enabling them to be explicit about their specific needs and (2) enables a care environment and service system to provide real-time interventions adapted to the patients' profiles and their changing status/needs. Advances in technologies have made this feasible. Next to the modified ICU room [20] and the conversational agent [21] described in the introduction, the intelligent ICU concept enables autonomous monitoring of patients and the ICU environment over time by using pervasive sensing [75]. Previously we pointed out that there were too few technologies aiming at improving patients' environment despite its significant influence on patients. Luetz et al [20] emphasized the potential of environment-related innovations, arguing that the ICU should be considered as a treatment tool. In order to design a healing environment, future technology-based interventions should reinforce the main ingredients of patient-centered care: context-awareness, personalization, and adaptability.

Study Limitations

This study has some limitations. First, this review was conducted between 2015 and 2020 to focus on state-of-the-art technologies used for preventing and reducing delirium. Therefore, some technologies that might have been introduced in the studies published before 2015 and have not been studied since then were not included in this review. Second, the search strategy made use of the combination of keywords describing different types of technologies and not the keyword "technology" as this was not widely used in literature on delirium prevention and reduction. Although we tried to cover all the existing technology-related keywords, this search strategy might have left out some very rare types of technologies. Third, the 7 pathways were made based on the included studies only. Therefore, some approaches and strategies from the studies, which did not meet our criteria, were not included, such as approaches and strategies related to pharmacological interventions or delirium detection only.

Conclusions

In this review, we provided an overview of technology-based interventions and proposed the 7 pathways to delirium prevention based on evidence-based studies. These insights can be considered as starting points for transforming ICUs into a healing environment, which might be well one of the most powerful nonpharmacological technology-based interventions for preventing delirium. Further research should generate a more in-depth and complete understanding of the key components of a healing environment for patients and on designing and developing technologies that can actualize it.

Acknowledgments

We would like to thank Thomas Falck, MSc (Philips Research) and Dr Ir Hanneke Becht (University of Twente) for helping with the development of the search strategy, Prof Dr Sophia de Rooij (University Medical Center Groningen) for advising on the formulation of the search scope, and Meyke Roosink (Philips) and Benedetta Lusi (University of Twente) for participating in the clustering session.

Authors' Contributions

CMK contributed to planning, screening, reading full text, and acquiring/analyzing/interpreting data. EMvdH contributed to screening, acquiring data, reading full text, and revising the results of data analysis/interpretation. TJLvR and GJV reviewed the results of data analysis/interpretation. GDSL conceptualized and reviewed the results of data analysis/interpretation. All authors contributed to drafting the manuscript or revising it critically for important intellectual content. All authors read and confirmed the final version of the manuscript and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of the work are appropriately investigated and resolved.

Conflicts of Interest

This study is part of the Digital Nature project that received funding from the Top Technology Twente Connecting Industry program (TKI Topsector HTSM), which is partially funded (paid to institution) by Philips. CMK is employed by the University of Twente through this fund. EMvdH is employed fully by Philips. The remaining authors do not have a conflict of interest to declare.

Multimedia Appendix 1
Search strategy.
[PDF File (Adobe PDF File), 74 KB - jmir_v23i8e26079_app1.pdf]

Multimedia Appendix 2
Quality assessment.
[PDF File (Adobe PDF File), 147 KB - jmir_v23i8e26079_app2.pdf]

Multimedia Appendix 3
Fourteen strategies used in the included studies.
[PDF File (Adobe PDF File), 229 KB - jmir_v23i8e26079_app3.pdf]

References

1. Morandi A, Jackson JC, Ely EW. Delirium in the intensive care unit. Int Rev Psychiatry 2009 Feb;21(1):43-58. [doi: 10.1080/09540260802675296] [Medline: 19219712]
2. Page VJ, Navarange S, Gama S, McAuley DF. Routine delirium monitoring in a UK critical care unit. Crit Care 2009;13(1):R16 [FREE Full text] [doi: 10.1186/cc7714] [Medline: 19203391]
3. Ely EW, Inouye SK, Bernard GR, Gordon S, Francis J, May L, et al. Delirium in mechanically ventilated patients: validity and reliability of the confusion assessment method for the intensive care unit (CAM-ICU). JAMA 2001 Dec 05;286(21):2703-2710. [doi: 10.1001/jama.286.21.2703] [Medline: 11730446]
4. Spronk PE, Riekerk B, Hofhuis J, Rommes JH. Occurrence of delirium is severely underestimated in the ICU during daily care. Intensive Care Med 2009 Jul;35(7):1276-1280 [FREE Full text] [doi: 10.1007/s00134-009-1466-8] [Medline: 19350214]
5. Faria RDS, Moreno R. Delirium in intensive care: an under-diagnosed reality. Rev Bras Ter Intensiva 2013;25(2):137-147 [FREE Full text] [doi: 10.5935/0103-507X.20130025] [Medline: 23917979]
6. Salluh JI, Soares M, Teles JM, Ceraso D, Raimondi N, Nava VS, Delirium Epidemiology in Critical Care Study Group. Delirium epidemiology in critical care (DECCA): an international study. Crit Care 2010;14(6):R210 [FREE Full text] [doi: 10.1186/cc9333] [Medline: 21092264]
7. Girard TD, Jackson JC, Pandharipande PP, Pun BT, Thompson JL, Shintani AK, et al. Delirium as a predictor of long-term cognitive impairment in survivors of critical illness. Critical Care Medicine 2010;38(7):1513-1520. [doi: 10.1097/ccm.0b013e3181e47be1]
8. Ely EW, Shintani A, Truman B, Speroff T, Gordon SM, Harrell FE, et al. Delirium as a predictor of mortality in mechanically ventilated patients in the intensive care unit. JAMA 2004 Apr 14;291(14):1753-1762. [doi: 10.1001/jama.291.14.1753] [Medline: 15082703]
9. Vasilevskis EE, Chandrasekhar R, Holtze CH, Graves J, Speroff T, Girard TD, et al. The Cost of ICU Delirium and Coma in the Intensive Care Unit Patient. Medical Care 2018;56(10):890-897. [doi: 10.1097/mlr.0000000000000975]
10. Rubin FH, Bellon J, Bilderback A, Urda K, Inouye SK. Effect of the Hospital Elder Life Program on Risk of 30-Day Readmission. J Am Geriatr Soc 2018 Jan;66(1):145-149 [FREE Full text] [doi: 10.1111/jgs.15132] [Medline: 29086425]
11. Zamoscik K, Godbold R, Freeman P. Intensive care nurses' experiences and perceptions of delirium and delirium care. Intensive Crit Care Nurs 2017 Jun;40:94-100. [doi: 10.1016/j.iccn.2017.01.003] [Medline: 28259522]

12. Leslie DL, Marcantonio ER, Zhang Y, Leo-Summers L, Inouye SK. One-year health care costs associated with delirium in the elderly population. Arch Intern Med 2008 Jan 14;168(1):27-32 [FREE Full text] [doi: 10.1001/archinternmed.2007.4] [Medline: 18195192]
13. Inouye SK, Westendorp RG, Saczynski JS. Delirium in elderly people. The Lancet 2014 Mar;383(9920):911-922. [doi: 10.1016/s0140-6736(13)60688-1]
14. Inouye SK, Bogardus ST, Charpentier PA, Leo-Summers L, Acampora D, Holford TR, et al. A Multicomponent Intervention to Prevent Delirium in Hospitalized Older Patients. N Engl J Med 1999 Mar 04;340(9):669-676. [doi: 10.1056/nejm199903043400901]
15. Devlin J, Skrobik Y, Gélinas C, Needham D, Slooter A, Pandharipande P, et al. Clinical Practice Guidelines for the Prevention and Management of Pain, Agitation/Sedation, Delirium, Immobility, and Sleep Disruption in Adult Patients in the ICU. Crit Care Med 2018 Sep;46(9):e825-e873. [doi: 10.1097/CCM.0000000000003299] [Medline: 30113379]
16. Marra A, Ely EW, Pandharipande PP, Patel MB. The ABCDEF Bundle in Critical Care. Crit Care Clin 2017 Apr;33(2):225-243 [FREE Full text] [doi: 10.1016/j.ccc.2016.12.005] [Medline: 28284292]
17. DAS-Taskforce 2015, Baron R, Binder A, Biniek R, Braune S, Buerkle H, et al. Evidence and consensus based guideline for the management of delirium, analgesia, and sedation in intensive care medicine. Revision 2015 (DAS-Guideline 2015) - short version. Ger Med Sci 2015;13:Doc19 [FREE Full text] [doi: 10.3205/000223] [Medline: 26609286]
18. Trogrlić Z, van der Jagt M, Bakker J, Balas MC, Ely EW, van der Voort PHJ, et al. A systematic review of implementation strategies for assessment, prevention, and management of ICU delirium and their effect on clinical outcomes. Crit Care 2015 Apr 09;19:157 [FREE Full text] [doi: 10.1186/s13054-015-0886-9] [Medline: 25888230]
19. van de Pol I, van Iterson M, Maaskant J. Effect of nocturnal sound reduction on the incidence of delirium in intensive care unit patients: An interrupted time series analysis. Intensive Crit Care Nurs 2017 Aug;41:18-25. [doi: 10.1016/j.iccn.2017.01.008] [Medline: 28351551]
20. Luetz A, Grunow JJ, Mörgeli R, Rosenthal M, Weber-Carstens S, Weiss B, et al. Innovative ICU Solutions to Prevent and Reduce Delirium and Post-Intensive Care Unit Syndrome. Semin Respir Crit Care Med 2019 Oct;40(5):673-686. [doi: 10.1055/s-0039-1698404] [Medline: 31826268]
21. Bott N, Wexler S, Drury L, Pollak C, Wang V, Scher K, et al. A Protocol-Driven, Bedside Digital Conversational Agent to Support Nurse Teams and Mitigate Risks of Hospitalization in Older Adults: Case Control Pre-Post Study. J Med Internet Res 2019 Oct 17;21(10):e13440. [doi: 10.2196/13440] [Medline: 31625949]
22. Kang J, Lee M, Ko H, Kim S, Yun S, Jeong Y, et al. Effect of nonpharmacological interventions for the prevention of delirium in the intensive care unit: A systematic review and meta-analysis. J Crit Care 2018 Dec;48:372-384. [doi: 10.1016/j.jcrc.2018.09.032] [Medline: 30300863]
23. Rivosecchi RM, Smithburger PL, Svec S, Campbell S, Kane-Gill SL. Nonpharmacological interventions to prevent delirium: an evidence-based systematic review. Crit Care Nurse 2015 Feb;35(1):39-50. [doi: 10.4037/ccn2015423] [Medline: 25639576]
24. Rivosecchi RM, Kane-Gill SL, Svec S, Campbell S, Smithburger PL. The implementation of a nonpharmacologic protocol to prevent intensive care delirium. J Crit Care 2016 Feb;31(1):206-211. [doi: 10.1016/j.jcrc.2015.09.031] [Medline: 26596509]
25. Bannon L, McGaughey J, Verghis R, Clarke M, McAuley DF, Blackwood B. The effectiveness of non-pharmacological interventions in reducing the incidence and duration of delirium in critically ill patients: a systematic review and meta-analysis. Intensive Care Med 2019 Jan;45(1):1-12. [doi: 10.1007/s00134-018-5452-x] [Medline: 30506354]
26. Martinez F, Tobar C, Hill N. Preventing delirium: should non-pharmacological, multicomponent interventions be used? A systematic review and meta-analysis of the literature. Age Ageing 2015 Mar;44(2):196-204. [doi: 10.1093/ageing/afu173] [Medline: 25424450]
27. Reston JT, Schoelles KM. In-facility delirium prevention programs as a patient safety strategy: a systematic review. Ann Intern Med 2013 Mar 05;158(5 Pt 2):375-380 [FREE Full text] [doi: 10.7326/0003-4819-158-5-201303051-00003] [Medline: 23460093]
28. Abraha I, Trotta F, Rimland JM, Cruz-Jentoft A, Lozano-Montoya I, Soiza RL, et al. Efficacy of Non-Pharmacological Interventions to Prevent and Treat Delirium in Older Patients: A Systematic Overview. The SENATOR project ONTOP Series. PLoS One 2015;10(6):e0123090 [FREE Full text] [doi: 10.1371/journal.pone.0123090] [Medline: 26062023]
29. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
30. Leentjens AF, Schieveld JN, Leonard M, Lousberg R, Verhey FR, Meagher DJ. A comparison of the phenomenology of pediatric, adult, and geriatric delirium. J Psychosom Res 2008 Feb;64(2):219-223. [doi: 10.1016/j.jpsychores.2007.11.003] [Medline: 18222136]
31. Holly C, Porter S, Echevarria M, Dreker M, Ruzehaji S. CE: Original Research: Recognizing Delirium in Hospitalized Children: A Systematic Review of the Evidence on Risk Factors and Characteristics. Am J Nurs 2018 Apr;118(4):24-36. [doi: 10.1097/01.NAJ.0000532069.55339.f9] [Medline: 29543606]

32. Pandharipande P, Cotton BA, Shintani A, Thompson J, Costabile S, Truman Pun B, et al. Motoric subtypes of delirium in mechanically ventilated surgical and trauma intensive care unit patients. Intensive Care Med 2007 Aug 18;33(10):1860-1860. [doi: 10.1007/s00134-007-0846-1]
33. Byun S, Song S, Kim J, Ryu T, Jeong M, Kim E. Mother's recorded voice on emergence can decrease postoperative emergence delirium from general anaesthesia in paediatric patients: a prospective randomised controlled trial. Br J Anaesth 2018 Aug;121(2):483-489 [FREE Full text] [doi: 10.1016/j.bja.2018.01.042] [Medline: 30032889]
34. Kim H, Jung SM, Yu H, Park S. Video Distraction and Parental Presence for the Management of Preoperative Anxiety and Postoperative Behavioral Disturbance in Children. Anesthesia & Analgesia 2015;121(3):778-784. [doi: 10.1213/ane.0000000000000839]
35. Rodriguez ST, Jang O, Hernandez JM, George AJ, Caruso TJ, Simons LE. Varying screen size for passive video distraction during induction of anesthesia in low-risk children: A pilot randomized controlled trial. Paediatr Anaesth 2019 Jun;29(6):648-655. [doi: 10.1111/pan.13636] [Medline: 30916447]
36. Dwairej DA, Obeidat HM, Aloweidi AS. Video game distraction and anesthesia mask practice reduces children's preoperative anxiety: A randomized clinical trial. J Spec Pediatr Nurs 2020 Jan;25(1):e12272. [doi: 10.1111/jspn.12272] [Medline: 31576651]
37. Eijlers R, Dierckx B, Staals LM, Berghmans JM, van der Schroeff MP, Strabbing EM, et al. Virtual reality exposure before elective day care surgery to reduce anxiety and pain in children: A randomised controlled trial. Eur J Anaesthesiol 2019 Oct;36(10):728-737 [FREE Full text] [doi: 10.1097/EJA.0000000000001059] [Medline: 31356373]
38. Ryu J, Oh A, Yoo H, Kim J, Park J, Han S. The effect of an immersive virtual reality tour of the operating theater on emergence delirium in children undergoing general anesthesia: A randomized controlled trial. Paediatr Anaesth 2019 Jan;29(1):98-105. [doi: 10.1111/pan.13535] [Medline: 30365231]
39. Lee C, Lee C, Hsu M, Lai C, Sung Y, Lin C, et al. Effects of Music Intervention on State Anxiety and Physiological Indices in Patients Undergoing Mechanical Ventilation in the Intensive Care Unit. Biol Res Nurs 2017 Mar;19(2):137-144. [doi: 10.1177/1099800416669601] [Medline: 27655993]
40. Damshens MH, Sanie MS, Javadpour S, Khaef MA, Rastgarian A. The Role of Music on the Delirium in Traumatic Patients: A Case Study in the ICU of Peymanieh Hospital of Jahrom, Fars Province, Iran. Ambient Science 2018 Jul;5(Sp1):97-101. [doi: 10.21276/ambi.2018.05.sp1.ra11]
41. Simons KS, Laheij RJF, van den Boogaard M, Moviat MAM, Paling AJ, Polderman FN, et al. Dynamic light application therapy to reduce the incidence and duration of delirium in intensive-care patients: a randomised controlled trial. The Lancet Respiratory Medicine 2016 Mar;4(3):194-202. [doi: 10.1016/s2213-2600(16)00025-4]
42. Potharajaroen S, Tangwongchai S, Tayjasanant T, Thawitsri T, Anderson G, Maes M. Bright light and oxygen therapies decrease delirium risk in critically ill surgical patients by targeting sleep and acid-base disturbances. Psychiatry Res 2018 Mar;261:21-27. [doi: 10.1016/j.psychres.2017.12.046] [Medline: 29276990]
43. Waszynski CM, Milner KA, Staff I, Molony SL. Using simulated family presence to decrease agitation in older hospitalized delirious patients: A randomized controlled trial. Int J Nurs Stud 2018 Jan;77:154-161. [doi: 10.1016/j.ijnurstu.2017.09.018] [Medline: 29100197]
44. Demoule A, Carreira S, Lavault S, Pallanca O, Morawiec E, Mayaux J, et al. Impact of earplugs and eye mask on sleep in critically ill patients: a prospective randomized study. Crit Care 2017 Nov 21;21(1):284 [FREE Full text] [doi: 10.1186/s13054-017-1865-0] [Medline: 29157258]
45. Giraud K, Pontin M, Sharples LD, Fletcher P, Dalgleish T, Eden A, et al. Use of a Structured Mirrors Intervention Does Not Reduce Delirium Incidence But May Improve Factual Memory Encoding in Cardiac Surgical ICU Patients Aged Over 70 Years: A Pilot Time-Cluster Randomized Controlled Trial. Front Aging Neurosci 2016;8:228 [FREE Full text] [doi: 10.3389/fnagi.2016.00228] [Medline: 27733826]
46. Lin Y, Shen W, Liu Y, Wang Q, Chen Q, Fang Z, et al. Visual preconditioning reduces emergence delirium in children undergoing ophthalmic surgery: a randomised controlled trial. Br J Anaesth 2018 Aug;121(2):476-482 [FREE Full text] [doi: 10.1016/j.bja.2018.03.033] [Medline: 30032888]
47. Mitchell ML, Kean S, Rattray JE, Hull AM, Davis C, Murfield JE, et al. A family intervention to reduce delirium in hospitalised ICU patients: A feasibility randomised controlled trial. Intensive Crit Care Nurs 2017 Jun;40:77-84. [doi: 10.1016/j.iccn.2017.01.001] [Medline: 28254205]
48. Johnson K, Fleury J, McClain D. Music intervention to prevent delirium among older patients admitted to a trauma intensive care unit and a trauma orthopaedic unit. Intensive Crit Care Nurs 2018 Aug;47:7-14. [doi: 10.1016/j.iccn.2018.03.007] [Medline: 29735284]
49. Munro CL, Cairns P, Ji M, Calero K, Anderson WM, Liang Z. Delirium prevention in critically ill adults through an automated reorientation intervention - A pilot randomized controlled trial. Heart Lung 2017;46(4):234-238. [doi: 10.1016/j.hrtlng.2017.05.002] [Medline: 28606450]
50. Lee WJ, Lee MJ. The Effects of Video Information on Delirium, Anxiety, and Nursing Satisfaction of Heart Surgery Patients. IJBSBT 2017 Dec 31;8(6):175-184. [doi: 10.14257/ijbsbt.2016.8.6.17]

51. Arbabi M, Zebardast J, Noorbala AA, Mohamadi M, Rahimnia M, Larijani R. Efficacy of Liaison Education and Environmental Changes on Delirium Incidence in ICU. Arch Neurosci 2018 Apr 20;5(2):e56019. [doi: 10.5812/archneurosci.56019]
52. Gómez Tovar LO, Díaz Suárez L, Cortés Muñoz F. Cuidados de enfermería basados en evidencia y modelo de Betty Neuman, para controlar estresores del entorno que pueden ocasionar delirium en unidad de cuidados intensivos. eglobal 2016 Jan 09;15(1):49. [doi: 10.6018/eglobal.15.1.210841]
53. Sharda N, Mattoon E, Matters L, Prewitt J, McDonald S, Sloane R, et al. Bach to the Basics: Implementation and Impact of a Postoperative, Inpatient Personalized Music Program for Older Adults. J Perianesth Nurs 2019 Apr;34(2):347-353. [doi: 10.1016/j.jopan.2018.05.006] [Medline: 30205935]
54. Smonig R, Magalhaes E, Bouadma L, Andremont O, de Montmollin E, Essardy F, et al. Impact of natural light exposure on delirium burden in adult patients receiving invasive mechanical ventilation in the ICU: a prospective study. Ann Intensive Care 2019 Oct 17;9(1):120 [FREE Full text] [doi: 10.1186/s13613-019-0592-x] [Medline: 31624936]
55. Cheong CY, Tan JAQ, Foong Y, Koh HM, Chen DZY, Tan JJC, et al. Creative Music Therapy in an Acute Care Setting for Older Patients with Delirium and Dementia. Dement Geriatr Cogn Dis Extra 2016;6(2):268-275 [FREE Full text] [doi: 10.1159/000445883] [Medline: 27489560]
56. Garry J, Casey K, Cole TK, Regensburg A, McElroy C, Schneider E, et al. A pilot study of eye-tracking devices in intensive care. Surgery 2016 Mar;159(3):938-944. [doi: 10.1016/j.surg.2015.08.012] [Medline: 26361099]
57. Suvajdzic M, Bihorac A, Rashidi P, Ruppert M, Williams S, Ozrazgat-Baslanti T. Developing a Patient-Centered Virtual Reality Healthcare System To Prevent the Onset of Delirium in ICU Patients. 2019 Presented at: 2019 IEEE 7th International Conference on Serious Games and Applications for Health (SeGAH); 5-7 August; Kyoto, Japan. [doi: 10.1109/SeGAH.2019.8882442]
58. Estrup S, Kjer CKW, Poulsen LM, Gøgenur I, Mathiesen O. Delirium and effect of circadian light in the intensive care unit: a retrospective cohort study. Acta Anaesthesiol Scand 2018 Mar;62(3):367-375. [doi: 10.1111/aas.13037] [Medline: 29148046]
59. Zachary W, Kirupananthan A, Cotter S, Barbara GH, Cooke RC, Sipho M. The impact of Hospital Elder Life Program interventions, on 30-day readmission Rates of older hospitalized patients. Arch Gerontol Geriatr 2020;86:103963. [doi: 10.1016/j.archger.2019.103963] [Medline: 31733512]
60. Pustjens T, Schoutens A, Janssen L, Heesen W. Effect of dynamic light at the coronary care unit on the length of hospital stay and development of delirium: a retrospective cohort study. J Geriatr Cardiol 2018 Sep 28;15(9):567-573 [FREE Full text] [doi: 10.11909/j.issn.1671-5411.2018.09.006] [Medline: 30344540]
61. Ost LG. Applied relaxation: description of a coping technique and review of controlled studies. Behav Res Ther 1987;25(5):397-409. [doi: 10.1016/0005-7967(87)90017-9] [Medline: 3318800]
62. Hassenzahl M, Diefenbach S, Göritz A. Needs, affect, and interactive products – Facets of user experience. Interacting with Computers 2010 Sep;22(5):353-362. [doi: 10.1016/j.intcom.2010.04.002]
63. Russell JA, Mehrabian A. Evidence for a three-factor theory of emotions. Journal of Research in Personality 1977 Sep;11(3):273-294. [doi: 10.1016/0092-6566(77)90037-x]
64. Hu R, Hegadoren KM, Wang X, Jiang X. An investigation of light and sound levels on intensive care units in China. Aust Crit Care 2016 May;29(2):62-67. [doi: 10.1016/j.aucc.2015.08.001] [Medline: 26307553]
65. Leach J. Psychological factors in exceptional, extreme and torturous environments. Extrem Physiol Med 2016;5:7 [FREE Full text] [doi: 10.1186/s13728-016-0048-y] [Medline: 27257476]
66. Kordts B, Kopetz J, Henkel A, Scherader A, Jochems N. Requirements and interaction patters for a novel interaction device for patients in intensive care. i-com 2019;18(1):2019. [doi: 10.1515/icom-2019-0004]
67. Schulman-Marcus J, Mookherjee S, Rice L, Lyubarova R. New Approaches for the Treatment of Delirium: A Case for Robotic Pets. Am J Med 2019 Jul;132(7):781-782. [doi: 10.1016/j.amjmed.2018.12.039] [Medline: 30738728]
68. Özcan E, Birdja D, Simonse L, Struijs A. Alarm in the ICU! envisioning patient monitoring alarm management in future intensive care units. In: Pfannstiel M, Rasche C, editors. Service Design and Service Thinking in Healthcare and Hospital Management. Cham: Springer; 2019:421-446.
69. Lee C, Lai C, Sung Y, Lai MY, Lin C, Lin L. Comparing effects between music intervention and aromatherapy on anxiety of patients undergoing mechanical ventilation in the intensive care unit: a randomized controlled trial. Qual Life Res 2017 Jul;26(7):1819-1829. [doi: 10.1007/s11136-017-1525-5] [Medline: 28236262]
70. Gerber SM, Jeitziner M, Sänger SD, Knobel SEJ, Marchal-Crespo L, Müri RM, et al. Comparing the Relaxing Effects of Different Virtual Reality Environments in the Intensive Care Unit: Observational Study. JMIR Perioper Med 2019 Nov 15;2(2):e15579 [FREE Full text] [doi: 10.2196/15579] [Medline: 33393906]
71. Ludden GD, van Rompay TJ, Niedderer K, Tournier I. Environmental design for dementia care - towards more meaningful experiences through design. Maturitas 2019 Oct;128:10-16. [doi: 10.1016/j.maturitas.2019.06.011] [Medline: 31561816]
72. Greene SM, Tuzzio L, Cherkin D. A framework for making patient-centered care front and center. Perm J 2012;16(3):49-53 [FREE Full text] [Medline: 23012599]
73. Sharon T. Self-Tracking for Health and the Quantified Self: Re-Articulating Autonomy, Solidarity, and Authenticity in an Age of Personalized Healthcare. Philos. Technol 2016 Apr 18;30(1):93-121. [doi: 10.1007/s13347-016-0215-5]

74. Mead N, Bower P. Patient-centredness: a conceptual framework and review of the empirical literature. Soc Sci Med 2000 Oct;51(7):1087-1110. [doi: 10.1016/s0277-9536(00)00098-8] [Medline: 11005395]
75. Davoudi A, Malhotra KR, Shickel B, Siegel S, Williams S, Ruppert M, et al. Intelligent ICU for Autonomous Patient Monitoring Using Pervasive Sensing and Deep Learning. Sci Rep 2019 May 29;9(1):8020 [FREE Full text] [doi: 10.1038/s41598-019-44004-w] [Medline: 31142754]

Abbreviations

ICU: intensive care unit
VR: virtual reality

Edited by R Kukafka; submitted 27.11.20; peer-reviewed by MC Reade, YL Leung; comments to author 26.12.20; revised version received 12.02.21; accepted 24.05.21; published 26.08.21.

Please cite as:
Kim CM, van der Heide EM, van Rompay TJL, Verkerke GJ, Ludden GDS
Overview and Strategy Analysis of Technology-Based Nonpharmacological Interventions for In-Hospital Delirium Prevention and Reduction: Systematic Scoping Review
J Med Internet Res 2021;23(8):e26079
URL: https://www.jmir.org/2021/8/e26079
doi:10.2196/26079
PMID:34435955

©Chan Mi Kim, Esther M van der Heide, Thomas J L van Rompay, Gijsbertus J Verkerke, Geke D S Ludden. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Review

The Influence of Web-Based Tools on Maternal and Neonatal Outcomes in Pregnant Adolescents or Adolescent Mothers: Mixed Methods Systematic Review

Jania J Y Wu[1], MBBS; Nurulhuda Ahmad[2], MBBS; Miny Samuel[3], MSc, PhD; Susan Logan[2,4], MBChB, MD, FRCOG, FFSRH; Citra N Z Mattar[2,4], MBBS, MMED, MRANZCOG, PhD, FAMS

[1]Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
[2]Department of Obstetrics and Gynaecology, National University of Singapore, Singapore, Singapore
[3]Research Support Unit, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
[4]Department of Obstetrics and Gynaecology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

Corresponding Author:
Citra N Z Mattar, MBBS, MMED, MRANZCOG, PhD, FAMS
Department of Obstetrics and Gynaecology
National University of Singapore
1E Kent Ridge Rd
Singapore
Singapore
Phone: 65 67722672
Email: citramattar@nus.edu.sg

Abstract

Background: Pregnant adolescent women increasingly seek support during pregnancy and the puerperium through digital platforms instead of the traditional support system of family, friends, and the community. However, it is uncertain whether digital, web-based tools are reliable and effective in providing information to the user on a variety of topics such as fetal development, pregnancy outcomes, delivery, and breastfeeding to improve maternal and infant outcomes.

Objective: We aimed to identify web-based tools designed to promote knowledge, attitudes, and skills of pregnant adolescents or adolescent mothers and determine the efficacy of such web-based tools compared with conventional resources in promoting good pregnancy and infant outcomes.

Methods: A systematic search was conducted using Medline, Scopus, CINAHL, and PsycINFO for articles published from January 2004 to November 2020 to identify randomized trials and observational studies that evaluated digital, web-based platforms to deliver resources to pregnant adolescents. All articles written in the author's languages were included. Two authors independently reviewed abstracts and full-text articles for inclusion and assessed study quality. Risk of bias in each study was assessed using appropriate tools recommended by PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) and the Joanna Briggs Institute. We adopted a qualitative synthesis and presented the results in a narrative format due to the heterogenous nature of the studies.

Results: Seven articles met the inclusion criteria and were analyzed. The majority of the studies were graded to be of low to moderate risk for bias. The research methodologies represented were varied, ranging from randomized (n=1) and nonrandomized controlled trials (n=1) and prospective cohort studies (n=1) to mixed methods studies (n=1) and longitudinal surveys (n=3). Four studies included active web-based interventions, and 3 described exposure to web-based tools, including the use of social media and/or other internet content. Web-based tools positively influenced treatment-seeking intentions (intervention 17.1%, control 11.5%, P=.003) and actual treatment-seeking behavior for depression among postpartum adolescents (intervention 14.1%, control 6.5%, P<.001). In contrast, readily available information on the internet may leave adolescents with increased anxiety. The critical difference lies in information curated by health care professionals specifically to address targeted concerns versus self-acquired data sourced from various websites.

Conclusions: Despite almost universal web use, few studies have used this platform for health promotion and disease prevention. Social media interventions or web-based tools have the potential to positively influence both maternal and infant outcomes in adolescent pregnancy, but there is a need for more well-conducted studies to demonstrate the effectiveness of these support programs. The vastness of the information available on the web limits the ability of health care professionals to monitor or control sources of information sought by patients. Thus, it is important to create professionally curated platforms to prevent or limit exposure to potentially misleading or harmful information on the internet while imparting useful knowledge to the user.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42020195854; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=195854

(J Med Internet Res 2021;23(8):e26786) doi:10.2196/26786

KEYWORDS

pregnancy in adolescence; teenagers; adolescents; pregnancy; postpartum; internet; digital health; digital media; new digital media; eHealth; social media; social network; communications media

Introduction

Pregnant adolescents are an especially vulnerable population. Despite the significant decline of adolescent pregnancies in recent decades [1], the World Health Organization estimates that 12 million girls aged 15 to 19 years give birth yearly in developing countries [2]. These adolescents are at increased risk of prenatal and perinatal complications including gestational hypertension, preterm delivery, low infant birth weight, and other neonatal complications [3-5]. Adolescent pregnancies are more prevalent in socioeconomically disadvantaged communities, on a background of disrupted family structures and limited educational opportunities [6]. Poor pregnancy outcomes are more frequent among socioeconomically disadvantaged adolescents, largely due to the complex social and cultural factors that result in lower or delayed maternal engagement with health care services [7,8].

The transition from child-free adolescence to motherhood is daunting. Traditionally, those who are pregnant or postpartum turn to their family, friends, and partners for support [9]. More recently, community-based and home-visit programs have also been established to support adolescent mothers. Home-visit programs may offer better outcomes in adolescents who are difficult to engage due to close bonds formed [10]. These programs aim to provide access to information, resources, and social support in order to maximize coping strategies and eventual reintegration into society [11,12]. The Teenage Pregnancy Strategy is an example of a successful, multicomponent intervention that has reduced adolescent conceptions and improved outcomes for adolescent mothers by providing support for mothers targeted at completion of education and securing appropriate housing [13].

With increasing access to technology, expectant mothers may seek pregnancy-related information or support from social media [14,15] or internet-based platforms [16,17] instead of traditional sources. Due to shifts in contemporary social structure, many women find themselves geographically and emotionally isolated from their support system of family and friends [14,15]. Alternative support systems on digital platforms provide opportunities for like-minded women to exchange experiences and gain social support. Web-based support systems are available regardless of time and location and allow for anonymity of use, reducing stigma and facilitating the discussion of sensitive topics [18]. Furthermore, the majority of expectant mothers perceive the internet to be a reliable source of information and access information on a large variety of topics including fetal development, delivery, and breastfeeding [16]. As such, information found on the internet has the potential to influence the mother’s decisions surrounding her pregnancy care [19].

While the definition of social media is dynamic and constantly evolving [20], in general, a social media site is an interactive online platform that facilitates the exchange of user-generated content [21,22]. In our study, we defined social media as any online platform that enables users to exchange content (eg, Facebook or Instagram), while internet content was defined as any online platform that does not facilitate content exchange among users (eg, websites or online reading materials) [21,22]. We defined web-based tools broadly as describing all services and technologies found on online platforms and consider both social media and internet content to be subsets of web-based tools [23]. Despite their convenience and easy access [24], these tools have limitations. Information found online may not be verified and may provide pregnant women with inaccurate, unreliable, or unsupported knowledge [25]. A meta-analysis evaluating the quality of online health information found that 70% of the included studies concluded that information sources on the internet were of low quality [26] and often provided advice with limited or no scientific evidence [27-29]. Specifically regarding women’s health, inaccurate celebrity-based advice has been highlighted [30]. Low-quality pregnancy-related information may be harmful or conflicting [29] and is often not discussed with health care providers to clarify misconceptions [24,31], all of which have the potential to negatively influence pregnancy outcomes. The unregulated online community can also produce negative experiences for the naïve user [32]. As the use of social media or internet resources during pregnancy is a relatively recent phenomenon, there is an opportunity to explore the association between its use among at-risk adolescents and perinatal outcomes.

The aims of this systematic review were to assess the impact of web-based tools used by pregnant adolescents or adolescent mothers on maternal and infant outcomes to compare these to conventional resources and critically appraise the evidence from relevant quantitative and qualitative studies. The research questions addressed were as follows:

- What types of available web-based tools are designed to promote knowledge, attitudes, and skills of pregnant adolescents or adolescent mothers?
- How effective are these web-based tools in promoting good pregnancy and infant outcomes compared with conventional resources?

Methods

Search Strategy

The protocol was registered with the International Prospective Register of Systematic Reviews (PROSPERO) [CRD42020195854]. We followed the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) guidelines [33] and conducted a systematic search of PubMed, Scopus, CINAHL, and PsycINFO electronic databases for articles published from January 2008 to November 2020. We restricted publications to the last 12 years to ensure studies promoted updated practices relating to pregnancy care. The search was initially conducted on May 15, 2020, and updated on December 5, 2020. The process of updating the search was guided by methods described by Bramer et al [34]. We also conducted a secondary search to identify studies published between 2004 and 2007 as the first concept of Web 2.0, which is defined as a network platform that spans across all devices, was introduced in 2004 [35]. No further studies met our inclusion criteria. Two librarians from the National University of Singapore Medical Library were consulted on the finalization of the search strategy. Search terms included “pregnant,” “adolescents,” “social media,” and “internet,” and the full search strategy can be found in Multimedia Appendix 1.

Selection of Studies

The study selection was conducted in two phases. During level 1 screening, two authors (JW, NA) independently screened all studies retrieved by electronic database searches based on key terms and resolved discrepancies by discussion with a third author until a list of studies for level 2 screening was agreed upon. During level 2 screening, the full texts of studies selected in level 1 were retrieved and independently reviewed by the same 2 authors to determine the eligibility of each study; discrepancies were again resolved by discussion with the third author until a final list of studies was agreed on. The reasons for exclusion were coded and recorded systematically.

Eligibility Criteria

All inclusion and exclusion criteria were defined a priori. We included both quantitative and qualitative studies that explored the association between the use of web-based tools by pregnant adolescents or adolescent mothers and maternal and neonatal outcomes. Studies that defined their population as adolescents or studied women aged 21 years and younger were included. Social media was defined as online platforms providing avenues to exchange content with other users (eg, Facebook, Instagram, Twitter, blogs, vlogs, forums, chatrooms) while internet content was defined as online platforms that did not provide direct opportunities to interact with other users (eg, websites, online reading materials, internet programs). Conventional tools developed to support adolescent mothers such as brochures, school-based counseling, community-based counseling, and group and personal counseling were the comparison of interest. Maternal outcomes measured were physical (nutrition, physical activity, breastfeeding practices, birth complications, and risky behaviors such as smoking and alcohol consumption) and psychosocial (mental health, depression, anxiety, loneliness and stress, self-esteem, birth preparedness, and parenting outcomes) factors. Infant or child outcomes included preterm birth, low birth weight, sudden infant death syndrome, and obesity. All outcomes collected met the inclusion criteria. We included studies with or without a comparison group that were relevant to answering our research questions and excluded reviews, abstracts, conference proceedings, letters, editorials, comments, opinions, and book chapters. We excluded studies that were not in English, Chinese, Malay, or French, the languages of the authors. Studies that only examined the use of social media or internet content with no quantitative or qualitative outcomes were also excluded.

Data Extraction

We extracted relevant evidence using a standard proforma including study design, settings, observational period, sample size, participant characteristics, description of intervention and comparison tools, maternal outcomes, infant or child outcomes, adjusted factors, findings, and limitations.

Risk of Bias Assessment

We conducted a mixed methods systematic review to assess the relevant studies using various critical appraisal tools that are validated and widely used. Qualitative studies [36] and nonrandomized controlled trials [37] were evaluated using the Joanna Briggs Institute instrument, which appraises articles as “included,” “excluded,” or “seek further info” based on 10- and 9-question checklists, respectively. Cohort or case-control studies were evaluated using the Newcastle-Ottawa Quality Assessment Scale [38], which rates articles according to selection, comparability, and exposure categories using a star system. Randomized controlled trials (RCT) were evaluated using the Cochrane Risk of Bias 2 describing a “fixed set of domains of bias, focusing on different aspects of trial design, conduct, and reporting” [39]. It appraises articles as low risk of bias, some concerns, and high risk of bias. Studies using mixed methodologies were evaluated using the Mixed Methods Appraisal Tool [40] based on a qualitative and quantitative scoring system. Four authors (JW, NA, SL, CM) independently and critically appraised each study for quality and potential bias. We resolved discrepancies by discussion until a consensus was reached.

Synthesis of Results

We categorized included studies into 2 groups to answer our research questions. To explore the variety and effectiveness of web-based tools in promoting knowledge, we analyzed and compared all 7 studies [41-47]. To determine their influence on pregnancy outcomes compared to conventional resources, we assessed 2 studies [41,42]. The studies analyzed were too heterogeneous to enable a formal meta-analysis, thus we adopted a qualitative synthesis and presented the results in a narrative format.

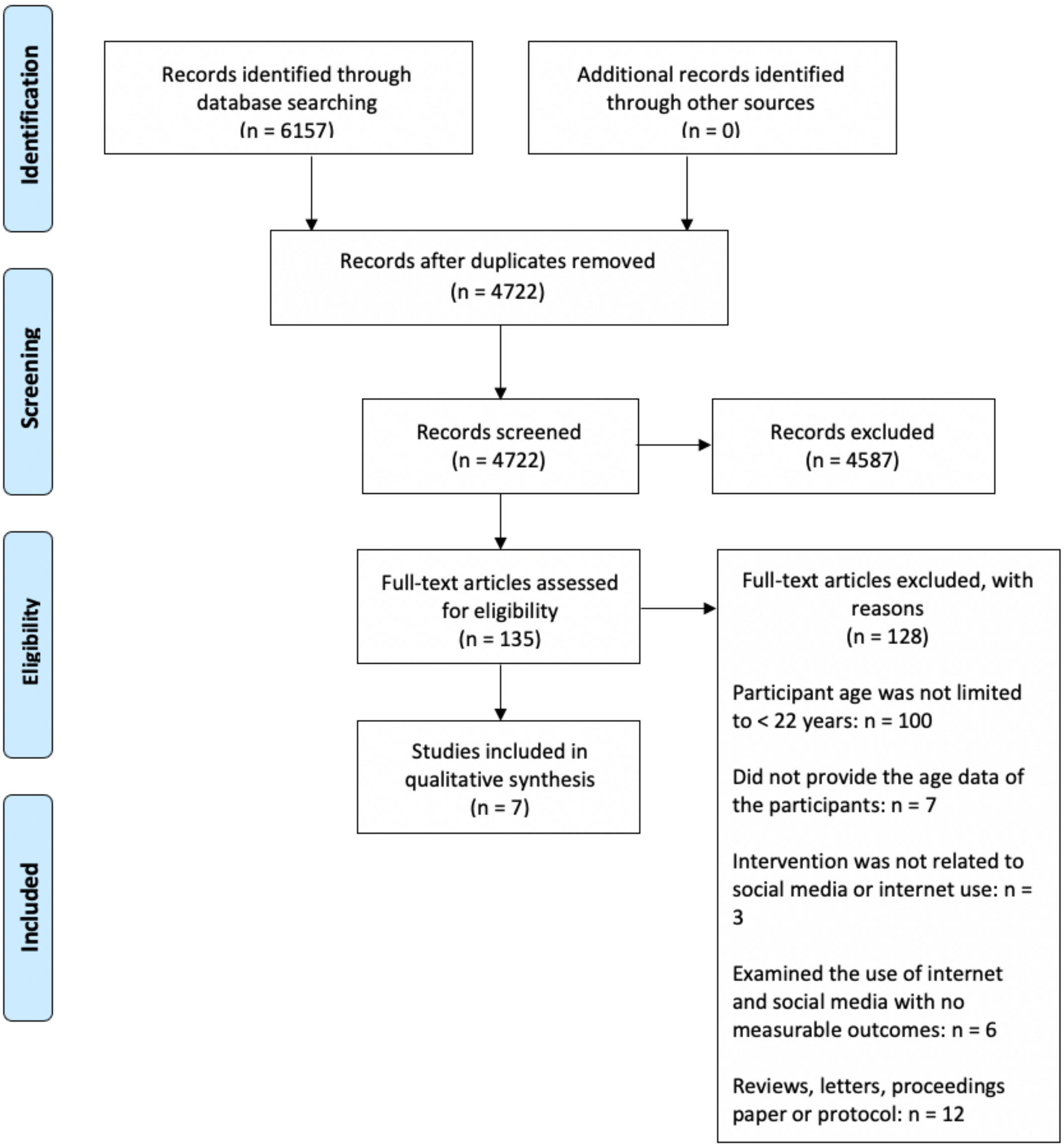
Results

Search Results

The search in the various medical databases (Medline [using PubMed platform], Scopus, CINAHL [using EbscoHost platform], and PsycINFO) for articles published from January

2008 to November 2020 yielded a total of 6157 records. After removing the duplicates, 4722 records were eligible for further screening. After 2 reviewers screened all titles and abstracts, 135 records met the inclusion criteria and were eligible for further full-text screening. Following the full text review, we excluded 128 articles based on the predefined exclusion criteria, and 7 articles were finally included in the systematic review for analysis (Figure 1).

Figure 1. PRISMA flow diagram.


Identification
Screening
Eligibility
Included
Records identified through database searching (n = 6157)
Additional records identified through other sources (n = 0)
Records after duplicates removed (n = 4722)
Records screened (n = 4722)
Records excluded (n = 4587)
Full-text articles assessed for eligibility (n = 135)
Full-text articles excluded, with reasons (n = 128)
Participant age was not limited to < 22 years: n = 100
Did not provide the age data of the participants: n = 7
Intervention was not related to social media or internet use: n = 3
Examined the use of internet and social media with no measurable outcomes: n = 6
Reviews, letters, proceedings paper or protocol: n = 12
Studies included in qualitative synthesis (n = 7)

Study Characteristics

Characteristics of the 7 studies are summarized in Table 1. Six were conducted in the United States [41-44,46,47], and one was conducted in Western Australia [45]. The research methodologies represented were varied—the most common study design was qualitative (n=3) [43,45,46], while the remaining studies were an RCT [41], a non-RCT [47], a prospective cohort study [42], and a mixed methods study [44]. Study sizes ranged from 7 [43,45] to 292 participants [42]. Despite limiting our inclusion criteria to adolescents, the participant age range was large, from age 13 years [42] to a participant aged 22 years, who was included as she was considered an adolescent by the authors [42,46]. Most participants were defined as adolescents in 6 studies [41,42,44-47], while 1 study defined participants as first-time mothers [43]. Participants were currently pregnant (n=1) [44], already mothers (n=5) [41-43,45,47], or both (n=1) [46]. Various methods were used to encourage participation: 5 studies recruited participants from established institutions such as prenatal clinics or organizations offering services for adolescents [41,42,44,46,47], while the remaining 2 studies recruited participants by posting public advertisements or through personal and professional contacts [43,45]. Two studies included a control group, which received conventional care [41,42].

Table 1. Descriptive characteristics.

Author, country	Study design	Participants	Intervention/exposure	Control group (if any)	Outcomes evaluated	Key findings
Hudson et al [41], US	Quantitative; RCT[a]	Adolescents: 16-21 years (mean 18.3 [SD 1.7] years); 1-week postpartum; single, low-income, African Americans	n=15; NMN[b] website: internet-based education resource, discussion forum, direct email contact with nurses (6 months)	n=19; usual care: hospital parenting instructions, parent's own resources	Maternal: • Mental health • Parenting outcomes • Birth complications Infant: • Health care use • Breastfeeding	The NMN website is well poised for nursing-driven social support intervention. The social support component was identified as a key strength with positive qualitative comments.
Logsdon et al [42], US	Quantitative; matched prospective cohort study	Adolescents: 13-21 years; up to 1-year postpartum; living in urban, suburban and rural counties, mixture of White (8.6%), Black (88.0%) and others (3.4%)	n=154 (mean 17.9 [SD 2.1] years); internet intervention website: internet-based education resources (2 weeks)	n=138 (mean 18.2 [SD 1.9] years); home visitation program	Maternal: • Mental health	The internet intervention was successful in changing attitudes, perceived control, intention to seek treatment, and actually seeking treatment. The intervention effect was equal in adolescents regardless of where they lived, but the impact on changing attitudes may be dose dependent.
Fleming et al [43], US	Qualitative	First-time mothers; mean 18-21 [SD 19.5] years); 6-8 weeks postpartum; single (85.7%), low-income	n=7; personal electronic media use: websites, internet blogs, internet chat rooms, online shows (duration not specified)	__[c]	Maternal: • Birth preparedness • Mental health	This study demonstrated adolescents' desire and need for clear, accurate, and easily accessible information about birthing. Providing credible electronic sources will educate the mothers and increase their confidence and birthing preparedness levels.
Vander Wyst et al [44], US	Mixed methods; non-RCT	Adolescents: 14-18 years; 12-28 weeks pregnant; low-income, mixture of Black (70%), Hispanic White (20%), and non-Hispanic White (10%)	n=10 (median 17.0 [IQR[d] 16.4, 17.7] years); social media intervention: private Facebook group with interactive activities and dissemination of health information (18 weeks)	n=12 (median 29.2 [IQR 23.7, 33.8] years); adult participants[e]	Maternal • Physical anthropometric data • Nutrition knowledge • Nutrition behavior • Physical activity • Attitudes and beliefs on prenatal health • Social support Infant • Birth weight • Gestational age • Breastfeeding	Poor diet quality persists among both adolescent and adult low-income pregnant women. Although social media-based education was well received by the participants, this did not result in significant changes in dietary intake and knowledge.
Nolan et al [45], WA[f]	Qualitative	Adolescents; 16-19 years; 3-17 months postpartum; single, living with parents (71.4%), extended relatives (14.3%), or partner/friend (14.3%)	n=7; personal social network site use: website that enables users to create public profiles and form relationships with other users (duration not specified)	—	Maternal: • Social support • Mental health • Parenting outcomes	The use of social network sites affords adolescent mothers access to tangible, informational, and emotional support. There is a potential role for midwives to use such platforms to provide additional social support.

Author, country	Study design	Participants	Intervention/exposure	Control group (if any)	Outcomes evaluated	Key findings
Rueda et al [46], US	Qualitative	Adolescents; 14-22 years; currently pregnant or mothers; single, living in residential foster care home, mixture of ethnic minorities: Hispanic (43.5%), Black (30.4%), Mixed race (10.9%)	n=13; personal electronic media use: social media websites, phone apps that facilitate communication between individuals (duration not specified)	—	Maternal: • Relationship with intimate partners • Mental health Infant: • Child protection	The use of technology among adolescent mothers living in foster homes is associated with multiple social issues. Technology should be included in various models of care to increase understanding between professionals and adolescents.
Logsdon et al [47], US	Quantitative; non-RCT	Adolescents; mean 16.8 years; mothers; single, students of a public school–based program for adolescent parents; mixture of African American (48.6%), White (34.1%), and others (17.3%)	n=138; internet intervention website: internet-based education resources (single class period)	—	Maternal • Mental health	The testing of a prototype website for adolescent mothers with postpartum depression showed promising results. Attitudes related to depression and seeking treatment improved after viewing the website.

[a] RCT: randomized controlled trial.

[b] NMN: New Mothers Network.

[c] Not applicable.

[d] IQR: interquartile range.

[e] Control group (adult participants) is not relevant to answering the research question.

[f] WA: Western Australia.

Risk of Bias Within Studies

Table 2 illustrates an overview of the studies' risk of bias. A detailed assessment of the relevant studies using various critical appraisal tools is found in Multimedia Appendix 2, as the criteria for each study differed by study design. All 7 studies were included, and 6 were judged to have a low risk of bias [41-46]. Of note, 2 out of 7 studies [42,47] were conducted by Logsdon et al [47]. The earlier study in 2013 was conducted as a pilot to test the prototype of a web-based intervention. Although included, it was rated as a poor-quality study due to the lack of a control group, the one-off nature of outcome sampling, lack of meaningful clinical outcome, and an inadequate description of statistical analysis used. This was subsequently followed by a more robust prospective cohort study in 2018 [42].

Table 2. Overview of the studies' risk of bias.

Author	Study design	Quality assessment instrument	Rating
Hudson et al [41]	Quantitative; RCT[a]	Cochrane Risk of Bias 2	Include; risk of bias: low
Logsdon et al [42]	Quantitative; matched prospective cohort study	Newcastle-Ottawa Quality Assessment Scale (Cohort studies)	Include; selection: ***; comparability: **; outcome: **; risk of bias: low
Vander Wyst et al [44]	Mixed methods; non-RCT	MMAT[b]	Include; risk of bias: low
Logsdon et al [47]	Quantitative; non-RCT	JBI[c] (quasi-experimental studies)	Include; risk of bias: moderate-high
Fleming et al [43]	Qualitative	JBI (qualitative research)	Include; risk of bias: low
Nolan et al [45]	Qualitative	JBI (qualitative research)	Include; risk of bias: low
Rueda et al [46]	Qualitative	JBI (qualitative research)	Include; risk of bias: low

[a] RCT: randomized controlled trial.

[b] MMAT: Mixed Methods Appraisal Tool.

[c] JBI: Joanna Briggs Institute.

Synthesis of Results

The study results are summarized in Tables 3 and 4. Adolescents in 4 studies [41,42,44,47] received an active intervention. Of these studies, 1 included adolescent–health care professional interactions [41], 2 described adolescent-adolescent interactions [41,44], and the remaining 2 studies described purely internet content [42,47]. The adolescents in the remaining 3 studies [43,45,46] were exposed to various web-based tools. Of these studies, the duration of exposure to the web-based tools was not known for all [43,45,46], and resources accessed by the adolescents include informative [43] and social media websites

[43,45,46]. Various outcomes were collected during and after the intervention, measured via self-reports, postintervention surveys, corroboration with medical records, validated tools (if available), and general anthropometric data. The outcomes are discussed as maternal and infant outcomes. We further categorized maternal outcomes into physical, mental well-being, parenting outcomes, and others.

Table 3. Synthesis of quantitative results.

Author, country	Statistically significant outcomes with intervention ($P<.05$, unless stated otherwise)	Non–statistically significant trends following intervention
Hudson et al [41], US	Assuming $\alpha=.10$, $P<.10$; • Intervention group had lower self-esteem than control group at 6 months; scale: RSE[a] • Intervention group had higher levels of perceived competence after 6 months; scale: PPS[b] • Intervention group had higher parenting satisfaction levels after 6 months; scale: WPBL-R[c] • ER[d] use reduced >50% in intervention group compared to control group (35.7% vs 70.6%); data collection: questionnaire • Intervention group was less likely to exclusively breastfeed compared to control group; data collection: questionnaire	Increasing: • Social support No differences in: • Depression symptoms • Loneliness • Perceived stress • Birth complications
Logsdon et al [42], US	• Intervention group had more positive attitudes toward seeking psychological help than the control group after 2 weeks; scale: ATSPH[e] • Intervention group had more positive perceived behavior control than the control group after 2 weeks; scale: HSDI[f] • Intervention group had greater intention to seek treatment for depression than the control group after 2 weeks; scale: MHI[g] • Intervention group had higher treatment seeking behavior for depression than the control group after 2 weeks; data collection: questionnaire	No differences in: • Depression symptoms • Stigma for receiving psychological help
Vander Wyst et al [44], US	• There was higher sugar intake in both groups after 18 weeks compared to baseline; data collection: 24-hour diet recall calculated via FPP[h] • There was a lower likelihood of adolescents cooking at home at baseline compared to adults; data collection: questionnaire • There was a lower likelihood of adolescents buying their own groceries at baseline and after 18 weeks compared to adults; data collection: questionnaire • Adolescents were less knowledgeable in nutrition (eg, identifying fiber rich food, recommended whole grain consumption, fruit, vegetable and fat intake) compared to adults at baseline and/or after 18 weeks; data collection method: questionnaire	No differences in: • Participant anthropometric data • Mean caloric consumption • Macronutrient distribution of food • Infant birth weight • Infant gestational age
Logsdon et al [47], US	• Adolescents had more positive attitudes toward seeking psychological help postintervention compared to baseline; scale: ATSPH	No differences in: • Mental health acceptability • Stigma for receiving psychological help

[a] RSE: Rosenberg Self-Esteem.

[b] PPS: How I Deal With Problems Regarding Care of My Baby.

[c] WPBL-R: What Being the Parent of a Baby is Like–Revised.

[d] ER: emergency room.

[e] ATSPH: Attitude Toward Seeking Psychological Help.

[f] HSDI: Health Self Determination Index.

[g] MHI: Mental Health Intention.

[h] FPP: Food Processor Program.

Table 4. Synthesis of qualitative results.

Study, country	Outcomes
Fleming et al [43], US	• Increased anxiety due to graphic media, birthing process, potential complications, and neonatal care • Birth preparedness: suboptimal birth preparedness due to fragmented, inconsistent, weakly linked, and poorly referenced information although a small subset of women developed improved or enhanced understanding • Social support: platform allowed connection with others and peer support
Vander Wyst et al [44], US	• Nutrition behavior: adolescents had improved attitudes toward nutrition with dietary changes (eg, limiting high fat fast food, increasing vegetable and fruit intake), motivated by time, convenience, and food preferences • Physical activity: adolescents had an increased tendency to exercise during pregnancy as they believed it to help with labor • Breastfeeding: adolescents tended to be less likely to breastfeed compared to adults • Social support: both adolescents and adults had both good and poor sources of social support
Nolan et al [45], WA[a]	• Increased social support and connectedness: participants had unlimited access to relationships, minimizing feelings of exclusion, and social isolation. They could maintain both old and new friendships. Social network sites provide valuable tangible, emotional, and informational support for adolescent mothers, contributing to mothers' social capital • Parental stress and anxiety: social network sites served as a medium for problem sharing and helped to reduce parental stress and anxiety. Drawbacks were the absence of adequate privacy controls and negative comments that could potentially threaten emotional well-being • Increased parenting confidence: peer support and positive affirmations significantly increased adolescents' confidence levels

Rueda et al [46], US	• Social media tools provided positive experiences in:
	  • Interacting with a potentially intimate partner
	  • Maintaining contact and fostering feelings of closeness with their child's father
	• Social media tools provided negative experiences in:
	  • Unwanted sexual solicitations
	  • Child protection, as meetings with strangers in offline spaces place both the adolescent and their children at risk
	  • Cyber abuse (eg, cyber bullying, stalking) of which adolescents were both victims and perpetuators
	  • Adverse emotional side effects fueled by jealousy and mistrust

[a] WA: Western Australia.

Maternal Outcomes

Maternal mental well-being was explored in 5 studies [41-43,45,47]. Hudson et al [41] conducted an RCT comparing participants who were exposed to the New Mothers Network website with those who received typical parenting instructions provided by the hospital. The New Mothers Network intervention provided parenting information through their electronic library and via communications with other mothers and nurses. The authors found that self-esteem levels measured via the Rosenberg Self-Esteem scale were significantly lower in the intervention group over the 6-month period (P=.04), although the authors found no clear explanation for this trend. In the 2 studies conducted by Logsdon et al [42,47], the earlier prospective pilot study [47] explored the efficacy of a website showcasing pictures and stories of other adolescent mothers' experiences, county and national mental health resources, and a frequently asked questions segment. The evidence from this study suggests that attitudes toward seeking psychological help, measured via the Attitude Toward Seeking Psychological Help scale, were significantly higher postintervention (P=.02). A similar result was observed in their prospective cohort study [42] of participants who received exposure to the website and home visits by volunteers as part of a home visitation program. Compared to the earlier pilot study [47], the intervention was adapted slightly to showcase video vignettes instead of pictures and stories based on user feedback. Mental health community resources and the frequently asked questions segment remained. The earlier results of improved attitudes toward seeking psychological help (P=.04) were successfully replicated, with additional data demonstrating that perceived behavior control (P=.007), intention to seek treatment for depression (P=.003), and seeking treatment for depression (P<.001) were significantly higher after 2 weeks in the intervention group compared to the controls. Data was measured using the Attitude Toward Seeking Psychological Help scale, Health Self Determination Index scale, Mental Health Intention scale, and a questionnaire on the outcomes of seeking treatment, respectively. Fleming et al [43] conducted qualitative interviews of participants who had prepared for childbirth through a variety of web-based tools including websites, blogs, chatrooms, and mass media. The findings suggested that access to electronic media did not necessarily prepare adolescents for pregnancy and childbirth but instead increased anxiety levels. Nolan et al [45] also conducted qualitative interviews on participants who used social media and reported that communicating with other adolescent mothers via social network sites helped reduce loneliness, parental stress, and anxiety.

Only one study by Vander Wyst et al [44] explored maternal physical outcomes. In this 18-week longitudinal study, participants were added into a private Facebook group where health information (pregnancy fitness, healthy recipes, nutrition, pregnancy fun facts, and stress management) were disseminated and interactive group activities were conducted. The authors compared adolescent and adult mothers and found dietary sugar distribution among both groups to be significantly increased after 18 weeks, with a greater increase in the adolescent group (7.4 [SD 0.2] vs 6.3 [SD 0.1] g/d, P=.005). Adolescent mothers were less likely to shop for groceries (postintervention; 14.0% vs 89.0%, *P*=.002) or cook at home (postintervention; 14.0% vs 67.0%, *P*=.054) compared to adult mothers and had significantly less knowledge regarding nutrition (identifying fiber rich foods, recommended daily consumption of wholegrains, fruit and vegetable varieties, and fat) before and after the intervention. There were no significant differences concerning anthropometric data, mean caloric consumption, or macronutrient distribution of food.

Parenting outcomes were explored in 3 studies [41,43,45]. Hudson et al [41] found that both intervention and control groups had significantly higher perceived competence (*P*<.01) and parenting satisfaction (*P*<.01) measured by the How I Deal With Problems Regarding Care of My Baby and evaluation subscale of the What Being the Parent of a Baby is Like–Revised, respectively. Fleming et al [43] concluded that although many young mothers had acquired knowledge on what to expect during childbirth, much of this was fragmented, inconsistent, weakly linked, poorly referenced, not always beneficial, and potentially a greater source of confusion. In contrast, Nolan et al [45] found that using social network sites increased young mothers’ confidence levels in parenting roles and with parenting strategies through positive affirmations, reassurance by other parents, and the collective sharing of experiences. Regarding social support to adolescent mothers, Hudson et al [41] observed a positive trend in social support levels following intervention supported by the qualitative comments submitted by participants. Nolan et al [45] described valuable tangible, emotional, and information support from social networks. Additionally, Rueda et al [46] explored the role of technology and social network sites on intimate relationships among adolescents living in residential foster homes. They determined these tools to be critical to the adolescent mothers’ ability to maintain or initiate intimate relationships with the fathers of their children or with new partners, facilitating both communication and in person meetings.

Infant Outcomes

Infant outcomes were studied less. None of the studies investigated putative associations between the use of web-based tools and premature birth, low birth weight, infant obesity, or sudden infant death syndrome. Hudson et al [41] found that emergency services use for postpartum problems in the first 6 months decreased significantly following intervention compared to no intervention, as 70.6% of mothers who did not receive intervention brought their child to the emergency room at least once compared to 35.7% of mothers who received the intervention (*P*=.052). In each group, there was one visit to the emergency room that was appropriate, with one infant hospitalized and a mother-infant pair treated for smoke inhalation. Vander Wyst et al [44] described adolescents as less likely to breastfeed compared to adults. Hudson et al [41] observed that adolescents in the intervention group were less likely to exclusively breastfeed compared to those who received usual care (*P*=.06 [assuming α=.10, *P*<.10]). Rueda et al [46] reported that residential foster home staff were concerned about adolescent mothers bringing their children to meetings arranged over social media with unknown partners, potentially placing their children at risk. While this topic was not discussed with the adolescent mothers, their dialogue suggests that they are aware of such risks. This belies an important knowledge and perception gap of program staff on how adolescent mothers use technology and social networks, as meeting strangers offline was either not as common as program staff believed or was regarded as shameful and not openly discussed by the adolescents.

Discussion

Principal Findings

Based on current available literature, there is much that needs to be explored concerning the potential benefits and harms of social media for pregnant and postnatal adolescents. This is despite most adolescent participants having grown up in the presence of advanced technology and with extensive social media use. We have systematically studied the impact of web-based tools used by adolescents on various maternal and infant outcomes to address our research questions. We found that available web-based tools include professionally curated programs (via websites or social media), readily available information on the internet, and personal use of various social media platforms.

While the limited evidence shows mixed and conflicting findings, we observe that web-based tools may be useful in improving mental health outcomes, positively influencing treatment-seeking intentions, and actual treatment-seeking behavior for depression among postpartum adolescents [42,47]. Conversely, readily available information on the internet may increase anxiety among adolescent recipients [43]. The difference appears to lie in the source of information, whether curated by health care professionals specifically to address common concerns of the target group and presented in a controlled setting or self-acquired data sourced from various websites initiated by the adolescent recipients themselves. RCTs using similar platforms, not limited to adolescent mothers, yield similar results, with postpartum mothers with depressive symptoms describing significantly improved parenting competence and decreased depression severity following social media interventions compared to unexposed mothers [48-51]. The presence of a professionally moderated program may be a key differentiating factor in mental health outcomes. Web-based tools may also be more influential on mental health outcomes due to the users’ relative anonymity compared to conventional care. A systematic review found that although online services did not significantly facilitate mental help-seeking behavior in youths, many youths regularly use online services and recommend them to peers as they are easily accessible, anonymous, and less stigmatizing [52].

Aside from mental health outcomes, there may be other benefits for adolescent mothers, including fewer emergency room visits (with increased knowledge and confidence in postpartum care) [41] and positive, albeit limited, dietary and behavioral changes [44]. Vulnerable populations like low-income adolescent mothers do not have ready access to health care professionals or services, but frequently have questions regarding childcare. These require rapid attention and resolution, resulting in unnecessary emergency room visits. Despite a lack of theoretical knowledge, structured interventions can motivate both adolescents and adults to seek healthier lifestyles. Furthermore, professional support via web-based tools has demonstrated positive impact on both infants [53,54] and older mothers [55-57] in the general prenatal population.

Preexisting social media platforms and internet-based tools play an integral role in an adolescent's life and are associated with positive and negative outcomes, particularly with maintaining beneficial friendships and fostering precarious and potentially dangerous relationships, respectively [45,46]. Despite some negative consequences, Sherman et al [32] found online adolescent pregnancy communities such as online support groups and discussion forums to be largely supportive and to serve an important role for those who use them. The authors noted that adolescents should choose their online community carefully to reduce additional psychological distress. Social media platforms may thus fill an unmet need to engage adolescent pregnant women, given the substantial use of social media globally [58].

Only 2 studies compared the efficacy of web-based tools and conventional methods in influencing maternal and neonatal outcomes in the adolescent population [41,42]. Mental health was the only common outcome that both studies evaluated, in which Logsdon et al [42] had a greater focus on depression compared to Hudson et al [41]. As such, no meaningful conclusions can be drawn to address our second research question.

A Singaporean study described increased risks of perinatal complications like anemia and preterm births and reduced likelihood of regular clinic attendance and sexually transmitted infection screening in a group of younger, vulnerable, predominantly Malay parturients with poorer access to prenatal care [59]. This highlights the importance of increasing accessibility to prenatal care and making a concerted effort to improve outreach [59]. Such young women are vulnerable to poor health literacy [60], closely related to eHealth literacy, which involves the use of digital technology [61], and are less likely to gain positive outcomes from internet searches [62]. eHealth interventions have proven effective in reaching out to populations with low levels of literacy [63], enabling health care professionals to facilitate behavioral change, personalize management, and improve education surrounding their health [64]. Aside from designing interventions to reach out to adolescents, it is also important to educate them and fill in the skill gap in eHealth literacy so they can access and evaluate health information accurately [65]. The use of professionally curated web-based tools incorporating online resources, communication services with health care professionals and peers, and social support may hold the key to reduce this important health inequality gap. However, larger age and culturally appropriate RCTs are necessary to validate the efficacy of these nontraditional methods.

Overall, our findings demonstrate a paucity of studies in this important aspect of managing adolescent pregnancies and highlight the effectiveness of web-based tools to reach pregnant and postpartum adolescents who may be more comfortable seeking help on online platforms. These tools allow health care professionals and policy makers to spread valuable pregnancy-related information to this vulnerable population in an age and peer acceptable way. Local or regional governments can potentially harness internet platforms and social media to drive public health policies. It may prove to be a more efficient allocation of resources, improving compliance to prenatal follow-ups and reducing pregnancy-related complications. Aside from antenatal care, governments can broaden their focus and cover topics pertaining to general women's health and, importantly, contraception use.

Strengths and Limitations

Our systematic review is timely as advanced technology and social media use are pervasive in modern societies, particularly among youth. Including both quantitative and qualitative studies in the review allowed for a broader interpretation of both statistical outcomes and adolescents' qualitative feedback.

Key limitations of this systematic review are the weaknesses inherent in the included studies and the lack of research in this important area, particularly in Asian societies, despite being the most connected globally. Notwithstanding the extensive internet and social media use among adolescents, few studies have been conducted to directly establish the relationship between technology use in pregnant adolescents or adolescent mothers and pregnancy outcomes. Across the 7 studies there was great variability in measurement scales and reported outcomes, contributing to the heterogeneity of our results, and making it challenging to draw well-grounded conclusions. For this reason, we were not able to conduct a meta-analysis and were only able to provide a descriptive narrative of the studies. These studies were based in the United States and Western Australia, and findings may not be directly applicable to more ethnically diverse populations given the cultural differences in family traditions, community infrastructure, and child-raising practices [66]. Babes Pregnancy, a charity based in Singapore [67], has used social media to support these vulnerable youth, although the impact of this program is not known.

Conclusion

The vastness of the information available on the web limits the ability of health care professionals to monitor or control sources of information sought by patients. It is important to create professionally curated platforms that patients can be directed to in order to prevent or limit exposure to potentially misleading or inaccurate information. Although our study is limited by the scope of the included studies, it is evident that web-based tools have the potential to improve outcomes in adolescent pregnancies. This review highlights the potential for web-based tools to target this vulnerable population, which is usually excluded from public health policies. The data presented can be highly informative to local health care policy makers. Overall, community-specific research is urgently needed to explore the potential of social media to guide novel interventions for this important vulnerable group.

Acknowledgments

We would like to thank the medical librarians Ms Wong Suei Nee and Ms Wong Huey Lih Winnifred from the National University of Singapore for their valuable feedback in the search strategy.

Authors' Contributions

JJYW conceptualized the research, conducted the database search, screened records, performed data extraction, critically appraised studies, analyzed the data, wrote the first draft of the paper, and reviewed the paper. NA conceptualized the research, screened articles, critically appraised studies, analyzed the data, and wrote and reviewed the paper. MS wrote and reviewed the paper. SL and CNZM conceptualized the research, critically appraised studies, and wrote and reviewed the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOC File , 27 KB - jmir_v23i8e26786_app1.doc]

Multimedia Appendix 2
Risk of bias.
[DOC File , 79 KB - jmir_v23i8e26786_app2.doc]

References

1. Sedgh G, Finer LB, Bankole A, Eilers MA, Singh S. Adolescent pregnancy, birth, and abortion rates across countries: levels and recent trends. J Adolesc Health 2015 Feb;56(2):223-230. [doi: 10.1016/j.jadohealth.2014.09.007] [Medline: 25620306]
2. Adolescent pregnancy. Geneva: World Health Organization; 2020 Jan 31. URL: https://www.who.int/news-room/fact-sheets/detail/adolescent-pregnancy [accessed 2021-07-26]
3. Cunnington AJ. What's so bad about teenage pregnancy? J Fam Plann Reprod Health Care 2001 Jan;27(1):36-41 [FREE Full text] [doi: 10.1783/147118901101194877] [Medline: 12457546]
4. Azevedo WFD, Diniz MB, Fonseca ESVB, Azevedo LMRD, Evangelista CB. Complications in adolescent pregnancy: systematic review of the literature. Einstein (Sao Paulo) 2015;13(4):618-626 [FREE Full text] [doi: 10.1590/S1679-45082015RW3127] [Medline: 26061075]
5. Chen X, Wen SW, Fleming N, Demissie K, Rhoads GG, Walker M. Teenage pregnancy and adverse birth outcomes: a large population based retrospective cohort study. Int J Epidemiol 2007 Apr;36(2):368-373. [doi: 10.1093/ije/dyl284] [Medline: 17213208]
6. Imamura M, Tucker J, Hannaford P, da Silva MO, Astin M, Wyness L, REPROSTAT 2 group. Factors associated with teenage pregnancy in the European Union countries: a systematic review. Eur J Public Health 2007 Dec;17(6):630-636. [doi: 10.1093/eurpub/ckm014] [Medline: 17387106]
7. Amjad S, MacDonald I, Chambers T, Osornio-Vargas A, Chandra S, Voaklander D, et al. Social determinants of health and adverse maternal and birth outcomes in adolescent pregnancies: a systematic review and meta-analysis. Paediatr Perinat Epidemiol 2019 Jan;33(1):88-99. [doi: 10.1111/ppe.12529] [Medline: 30516287]
8. Kapaya H, Mercer E, Boffey F, Jones G, Mitchell C, Anumba D. Deprivation and poor psychosocial support are key determinants of late antenatal presentation and poor fetal outcomes: a combined retrospective and prospective study. BMC Pregnancy Childbirth 2015 Nov 25;15:309 [FREE Full text] [doi: 10.1186/s12884-015-0753-3] [Medline: 26608259]
9. Bunting L, McAuley C. Research Review: teenage pregnancy and motherhood: the contribution of support. Child Fam Soc Work 2004 May;9(2):207-215. [doi: 10.1111/j.1365-2206.2004.00328.x]
10. Gomby DS, Culross PL, Behrman RE. Home visiting: recent program evaluations—analysis and recommendations. Future Child 1999;9(1):4-26. [Medline: 10414008]
11. Smithbattle L. Helping teen mothers succeed. J Sch Nurs 2006 Jun;22(3):130-135. [doi: 10.1177/10598405060220030201] [Medline: 16704281]
12. Culp AM, Culp RE, Blankemeyer M, Passmark L. Parent education home visitation program: adolescent and nonadolescent mother comparison after six months of intervention. Infant Ment Health J 1998;19(2):111-123. [doi: 10.1002/(SICI)1097-0355(199822)19:2<111::AID-IMHJ3>3.0.CO;2-R] [Medline: 12294463]
13. Skinner SR, Marino JL. England's teenage pregnancy strategy: a hard-won success. Lancet 2016 Aug 06;388(10044):538-540 [FREE Full text] [doi: 10.1016/S0140-6736(16)30589-X] [Medline: 27229191]
14. Holtz B, Smock A, Reyes-Gastelum D. Connected motherhood: social support for moms and moms-to-be on Facebook. Telemed J E Health 2015 May;21(5):415-421. [doi: 10.1089/tmj.2014.0118] [Medline: 25665177]

15. Baker B, Yang I. Social media as social support in pregnancy and the postpartum. Sex Reprod Healthc 2018 Oct;17:31-34. [doi: 10.1016/j.srhc.2018.05.003] [Medline: 30193717]
16. Bjelke M, Martinsson A, Lendahls L, Oscarsson M. Using the Internet as a source of information during pregnancy: a descriptive cross-sectional study in Sweden. Midwifery 2016 Sep;40:187-191. [doi: 10.1016/j.midw.2016.06.020] [Medline: 27450590]
17. Lupton D, Pedersen S, Thomas GM. Parenting and digital media: from the early web to contemporary digital society. Sociol Compass 2016 Aug 02;10(8):730-743. [doi: 10.1111/soc4.12398]
18. Herman J, Mock K, Blackwell D, Hulsey T. Use of a pregnancy support web site by low-income African American women. J Obstet Gynecol Neonatal Nurs 2005;34(6):713-720. [doi: 10.1177/0884217505282019] [Medline: 16282229]
19. Lagan BM, Sinclair M, Kernohan WG. Internet use in pregnancy informs women's decision making: a web-based survey. Birth 2010 Jun;37(2):106-115. [doi: 10.1111/j.1523-536X.2010.00390.x] [Medline: 20557533]
20. Aichner T, Grünfelder M, Maurer O, Jegeni D. Twenty-five years of social media: a review of social media applications and definitions from 1994 to 2019. Cyberpsychol Behav Soc Netw 2021 Apr;24(4):215-222 [FREE Full text] [doi: 10.1089/cyber.2020.0134] [Medline: 33847527]
21. Aichner T, Jacob F. Measuring the degree of corporate social media use. Int J Market Res 2015 Mar;57(2):257-276. [doi: 10.2501/ijmr-2015-018]
22. Kaplan AM, Haenlein M. Users of the world, unite! The challenges and opportunities of social media. Bus Horizons 2010 Jan;53(1):59-68. [doi: 10.1016/j.bushor.2009.09.003]
23. Boulos MNK, Maramba I, Wheeler S. Wikis, blogs and podcasts: a new generation of Web-based tools for virtual collaborative clinical practice and education. BMC Med Educ 2006;6:41 [FREE Full text] [doi: 10.1186/1472-6920-6-41] [Medline: 16911779]
24. Sayakhot P, Carolan-Olah M. Internet use by pregnant women seeking pregnancy-related information: a systematic review. BMC Pregnancy Childbirth 2016 Mar 28;16:65 [FREE Full text] [doi: 10.1186/s12884-016-0856-5] [Medline: 27021727]
25. Narasimhulu DM, Karakash S, Weedon J, Minkoff H. Patterns of internet use by pregnant women, and reliability of pregnancy-related searches. Matern Child Health J 2016 Dec;20(12):2502-2509. [doi: 10.1007/s10995-016-2075-0] [Medline: 27456311]
26. 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]
27. Chang T, Verma BA, Shull T, Moniz MH, Kohatsu L, Plegue MA, et al. Crowdsourcing and the accuracy of online information regarding weight gain in pregnancy: a descriptive study. J Med Internet Res 2016 Apr 07;18(4):e81 [FREE Full text] [doi: 10.2196/jmir.5138] [Medline: 27056465]
28. Kavlak O, Atan, Güleç D, Oztürk R, Atay N. Pregnant women's use of the internet in relation to their pregnancy in Izmir, Turkey. Inform Health Soc Care 2012 Dec;37(4):253-263. [doi: 10.3109/17538157.2012.710686] [Medline: 22958087]
29. Jesper E. Poor quality online information for pregnant women is a global problem: what can we do about it? BJOG 2015 Apr;122(5):740. [doi: 10.1111/1471-0528.13091] [Medline: 25315320]
30. Wiseman E. Jennifer Gunter: "Women are being told lies about their bodies". The Guardian. 2019 Sep 08. URL: https://www.theguardian.com/lifeandstyle/2019/sep/08/jennifer-gunter-gynaecologist-womens-health-bodies-myths-and-medicine [accessed 2021-07-26]
31. Larsson M. A descriptive study of the use of the Internet by women seeking pregnancy-related information. Midwifery 2009 Feb;25(1):14-20. [doi: 10.1016/j.midw.2007.01.010] [Medline: 17408822]
32. Sherman LE, Greenfield PM. Forging friendship, soliciting support: a mixed-method examination of message boards for pregnant teens and teen mothers. Comput Human Behav 2013 Jan;29(1):75-85. [doi: 10.1016/j.chb.2012.07.018]
33. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
34. Bramer W, Bain P. Updating search strategies for systematic reviews using EndNote. J Med Libr Assoc 2017 Jul;105(3):285-289 [FREE Full text] [doi: 10.5195/jmla.2017.183] [Medline: 28670219]
35. O'Reilly T. What is Web 2.0: design patterns and business models for the next generation of software. 2005 Sep 03. URL: https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html [accessed 2021-07-26]
36. Lockwood C, Munn Z, Porritt K. Qualitative research synthesis: methodological guidance for systematic reviewers utilizing meta-aggregation. Int J Evid Based Healthc 2015 Sep;13(3):179-187. [doi: 10.1097/XEB.0000000000000062] [Medline: 26262565]
37. Tufanaru C, Munn Z, Aromataris E, Campbell J, Hopp L. Chapter 3: Systematic reviews of effectiveness. In: Aromataris E, Munn Z, editors. JBI Manual for Evidence Synthesis. Adelaide: Joanna Briggs Institute; 2020.
38. Wells GA, Shea B, O'Connel D, Peterson J. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. 2009. URL: http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp [accessed 2021-07-26]
39. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]

40. Pluye P, Robert E, Cargo M, Bartlett G, O'Cathain A, Griffiths F. Proposal: a mixed methods appraisal tool for systematic mixed studies reviews. 2018. URL: http://mixedmethodsappraisaltoolpublic.pbworks.com/w/file/fetch/127916259/MMAT_2018_criteria-manual_2018-08-01_ENG.pdf [accessed 2020-11-09]
41. Hudson DB, Campbell-Grossman C, Hertzog M. Effects of an internet intervention on mothers' psychological, parenting, and health care utilization outcomes. Issues Compr Pediatr Nurs 2012;35(3-4):176-193. [doi: 10.3109/01460862.2012.734211] [Medline: 23140413]
42. Cynthia Logsdon M, Myers J, Rushton J, Gregg JL, Josephson AM, Davis DW, et al. Efficacy of an Internet-based depression intervention to improve rates of treatment in adolescent mothers. Arch Womens Ment Health 2018 Jun;21(3):273-285 [FREE Full text] [doi: 10.1007/s00737-017-0804-z] [Medline: 29260321]
43. Fleming SE, Vandermause R, Shaw M. First-time mothers preparing for birthing in an electronic world: internet and mobile phone technology. J Reprod Infant Psychol 2014 Mar 19;32(3):240-253. [doi: 10.1080/02646838.2014.886104]
44. Vander Wyst KB, Vercelli ME, O'Brien KO, Cooper EM, Pressman EK, Whisner CM. A social media intervention to improve nutrition knowledge and behaviors of low income, pregnant adolescents and adult women. PLoS One 2019;14(10):e0223120 [FREE Full text] [doi: 10.1371/journal.pone.0223120] [Medline: 31647852]
45. Nolan S, Hendricks J, Towell A. Social networking sites (SNS): exploring their uses and associated value for adolescent mothers in Western Australia in terms of social support provision and building social capital. Midwifery 2015 Sep;31(9):912-919. [doi: 10.1016/j.midw.2015.05.002] [Medline: 26026197]
46. Rueda HA, Brown ML, Geiger JM. Technology and dating among pregnant and parenting youth in residential foster care: a mixed qualitative approach comparing staff and adolescent perspectives. J Adolesc Res 2019 Jul 17;35(4):521-545. [doi: 10.1177/0743558419861087]
47. Logsdon MC, Barone M, Lynch T, Robertson A, Myers J, Morrison D, et al. Testing of a prototype Web based intervention for adolescent mothers on postpartum depression. Appl Nurs Res 2013 Aug;26(3):143-145. [doi: 10.1016/j.apnr.2013.01.005] [Medline: 23473677]
48. Boyd RC, Price J, Mogul M, Yates T, Guevara JP. Pilot RCT of a social media parenting intervention for postpartum mothers with depression symptoms. J Reprod Infant Psychol 2019 Jul;37(3):290-301. [doi: 10.1080/02646838.2018.1556788] [Medline: 30556428]
49. Forsell E, Bendix M, Holländare F, Szymanska von Schultz B, Nasiell J, Blomdahl-Wetterholm M, et al. Internet delivered cognitive behavior therapy for antenatal depression: a randomised controlled trial. J Affect Disord 2017 Oct 15;221:56-64 [FREE Full text] [doi: 10.1016/j.jad.2017.06.013] [Medline: 28628768]
50. Haga SM, Drozd F, Lisøy C, Wentzel-Larsen T, Slinning K. Mamma Mia: a randomized controlled trial of an internet-based intervention for perinatal depression. Psychol Med 2019 Aug;49(11):1850-1858 [FREE Full text] [doi: 10.1017/S0033291718002544] [Medline: 30191779]
51. Yang M, Jia G, Sun S, Ye C, Zhang R, Yu X. Effects of an online mindfulness intervention focusing on attention monitoring and acceptance in pregnant women: a randomized controlled trial. J Midwifery Womens Health 2019 Jan 29;64(1):68-77. [doi: 10.1111/jmwh.12944] [Medline: 30695166]
52. Kauer SD, Mangan C, Sanci L. Do online mental health services improve help-seeking for young people? A systematic review. J Med Internet Res 2014;16(3):e66 [FREE Full text] [doi: 10.2196/jmir.3103] [Medline: 24594922]
53. Huang MZ, Kuo S, Avery MD, Chen W, Lin K, Gau M. Evaluating effects of a prenatal web-based breastfeeding education programme in Taiwan. J Clin Nurs 2007 Aug;16(8):1571-1579 [FREE Full text] [doi: 10.1111/j.1365-2702.2006.01843.x] [Medline: 17655546]
54. Abuidhail J, Mrayan L, Jaradat D. Evaluating effects of prenatal web-based breastfeeding education for pregnant mothers in their third trimester of pregnancy: Prospective randomized control trial. Midwifery 2019 Feb;69:143-149. [doi: 10.1016/j.midw.2018.11.015] [Medline: 30513445]
55. Kuo S, Chen Y, Lin K, Lee T, Hsu C. Evaluating the effects of an Internet education programme on newborn care in Taiwan. J Clin Nurs 2009 Jun;18(11):1592-1601. [doi: 10.1111/j.1365-2702.2008.02732.x] [Medline: 19490296]
56. Smith K, Lanningham-Foster L, Welch A, Campbell C. Web-based behavioral intervention increases maternal exercise but does not prevent excessive gestational weight gain in previously sedentary women. J Phys Act Health 2016 Jun;13(6):587-593. [doi: 10.1123/jpah.2015-0219] [Medline: 26594820]
57. Bianchi CM, Mariotti F, Lluch A, Journet C, Stehr Y, Beaussier H, et al. Computer-based tailored dietary counselling improves the nutrient adequacy of the diet of French pregnant women: a randomised controlled trial. Br J Nutr 2019 Dec 18;123(2):220-231. [doi: 10.1017/s0007114519002617]
58. Twenge JM, Martin GN, Spitzberg BH. Trends in U.S. adolescents' media use, 1976–2016: the rise of digital media, the decline of TV, and the (near) demise of print. Psychol Popular Media Culture 2019 Oct;8(4):329-345. [doi: 10.1037/ppm0000203]
59. Tung JS, Kwek LK, Nadarajah S, Sulaiman S. Birth outcomes and successive pregnancy rates in adolescents attending a specialised antenatal clinic. Ann Acad Med Singap 2018 Jan;47(1):44-47 [FREE Full text] [Medline: 29493711]
60. MacLean L. A literature review of health literacy of adolescents during pregnancy. Nurs Womens Health 2020 Dec;24(6):431-439. [doi: 10.1016/j.nwh.2020.09.004] [Medline: 33157069]

61. Manganello J, Gerstner G, Pergolino K, Graham Y, Falisi A, Strogatz D. The relationship of health literacy with use of digital technology for health information: implications for public health practice. J Public Health Manag Pract 2016 Dec 15:380-387. [doi: 10.1097/PHH.0000000000000366] [Medline: 26672402]
62. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]
63. Jacobs RJ, Lou JQ, Ownby RL, Caballero J. A systematic review of eHealth interventions to improve health literacy. Health Informatics J 2014 Jun 10:81-98. [doi: 10.1177/1460458214534092] [Medline: 24916567]
64. Lotto M, Strieder AP, Ayala Aguirre PE, Oliveira TM, Andrade Moreira Machado MA, Rios D, et al. Parental-oriented educational mobile messages to aid in the control of early childhood caries in low socioeconomic children: a randomized controlled trial. J Dent 2020 Oct;101:103456 [FREE Full text] [doi: 10.1016/j.jdent.2020.103456] [Medline: 32827598]
65. Skopelja EN, Whipple EC, Richwine P. Reaching and teaching teens: adolescent health literacy and the internet. J Consum Heal Internet 2008 Jun 17;12(2):105-118. [doi: 10.1080/15398280802121406]
66. Forehand R, Kotchick BA. Cultural diversity: a wake-up call for parent training. Behav Ther 1996;27(2):187-206. [doi: 10.1016/s0005-7894(96)80014-1]
67. Babes: a helping hand for pregnant teens. URL: https://www.babes.org.sg/ [accessed 2021-07-26]

Abbreviations

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-analyses
PROSPERO: International Prospective Register of Systematic Reviews
RCT: randomized controlled trial

Edited by R Kukafka; submitted 27.12.20; peer-reviewed by M Lotto, E Said-Hung; comments to author 19.04.21; revised version received 12.06.21; accepted 05.07.21; published 26.08.21.

Please cite as:
Wu JJY, Ahmad N, Samuel M, Logan S, Mattar CNZ
The Influence of Web-Based Tools on Maternal and Neonatal Outcomes in Pregnant Adolescents or Adolescent Mothers: Mixed Methods Systematic Review
J Med Internet Res 2021;23(8):e26786
URL: https://www.jmir.org/2021/8/e26786
doi:10.2196/26786
PMID:34435961

©Jania J Y Wu, Nurulhuda Ahmad, Miny Samuel, Susan Logan, Citra N Z Mattar. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Review

Hospital Investment Decisions in Healthcare 4.0 Technologies: Scoping Review and Framework for Exploring Challenges, Trends, and Research Directions

Roberto Santiago Vassolo[1,2], DPhil; Alejandro Francisco Mac Cawley[2], DPhil; Guilherme Luz Tortorella[1,3,4], DPhil; Flavio Sanson Fogliatto[5], DPhil; Diego Tlapa[6], DPhil; Gopalakrishnan Narayanamurthy[7], DPhil

[1]IAE Business School, Universidad Austral, Pilar, Argentina
[2]Department of Industrial and Systems Engineering, Pontificia Universidad Católica de Chile, Santiago, Chile
[3]Department of Mechanical Engineering, University of Melbourne, Melbourne, Australia
[4]Universidade Federal de Santa Catarina, Florianopolis, Brazil
[5]Departamento de Engenharia de Produção, Universidade Federal do Rio Grande do Sul, Escola de Engenharia, Porto Alegre, Brazil
[6]Facultad de Ingeniería, Arquitectura y Diseño, Universidad Autónoma de Baja California - Campus Ensenada, Baja California, Mexico
[7]Department of Operations and Supply Chain Management, University of Liverpool Management School, Liverpool, United Kingdom

Corresponding Author:
Roberto Santiago Vassolo, DPhil
IAE Business School
Universidad Austral
Mariano Acosta s/n y Ruta Prov 8
Pilar, B1629WWA
Argentina
Phone: 54 2304481000
Email: rvassolo@iae.edu.ar

Abstract

Background: Alternative approaches to analyzing and evaluating health care investments in state-of-the-art technologies are being increasingly discussed in the literature, especially with the advent of Healthcare 4.0 (H4.0) technologies or eHealth. Such investments generally involve computer hardware and software that deal with the storage, retrieval, sharing, and use of health care information, data, and knowledge for communication and decision-making. Besides, the use of these technologies significantly increases when addressed in bundles. However, a structured and holistic approach to analyzing investments in H4.0 technologies is not available in the literature.

Objective: This study aims to analyze previous research related to the evaluation of H4.0 technologies in hospitals and characterize the most common investment approaches used. We propose a framework that organizes the research associated with hospitals' H4.0 technology investment decisions and suggest five main research directions on the topic.

Methods: To achieve our goal, we followed the standard procedure for scoping reviews. We performed a search in the Crossref, PubMed, Scopus, and Web of Science databases with the keywords *investment*, *health*, *industry 4.0*, *investment*, *health technology assessment*, *healthcare 4.0*, and *smart* in the title, abstract, and keywords of research papers. We retrieved 5701 publications from all the databases. After removing papers published before 2011 as well as duplicates and performing further screening, we were left with 244 articles, from which 33 were selected after in-depth analysis to compose the final publication portfolio.

Results: Our findings show the multidisciplinary nature of the research related to evaluating hospital investments in H4.0 technologies. We found that the most common investment approaches focused on cost analysis, single technology, and single decision-maker involvement, which dominate bundle analysis, H4.0 technology value considerations, and multiple decision-maker involvement.

Conclusions: Some of our findings were unexpected, given the interrelated nature of H4.0 technologies and their multidimensional impact. Owing to the absence of a more holistic approach to H4.0 technology investment decisions, we identified five promising research directions for the topic: development of economic valuation methodologies tailored for H4.0 technologies; accounting for technology interrelations in the form of bundles; accounting for uncertainties in the process of evaluating such technologies;

integration of administrative, medical, and patient perspectives into the evaluation process; and balancing and handling complexity in the decision-making process.

(J Med Internet Res 2021;23(8):e27571) doi:10.2196/27571

KEYWORDS

healthcare 4.0; scoping review; investments; real options; health technology assessment; technological bundles; decision-makers; hospital; public health; technology; health technology; smart technology; hospital management; health care investment; decision making; new technologies

Introduction

Background

How do health care organizations manage and determine their investment decisions in Industry 4.0 (known as Healthcare 4.0 [H4.0]) technologies? Having the right answer to this question is essential because the health care value chain is increasingly applying H4.0 technologies [1]. In addition, the rising demand for more efficient, qualified, and less expensive health services has motivated novel technological solutions [2]. Health care organizations have incorporated innovative technologies around the internet to facilitate and support more efficient and flexible processes, services, and products [3,4]. Such technologies started playing a pivotal role as enhancers of efficiency and quality in health care systems in the 1990s, culminating in what is currently known as *eHealth* [5]. Health care institutions extend the emerging principles and technologies belonging to the Industry 4.0 realm to health care as a continuous and disruptive process of innovation and transformation of the entire health care value chain [6].

The magnitude of the technological shift, the scope of activities affected, and their interrelationships expose health care decision-makers to large and complex investment decision problems [7,8]. The scope of activities encompasses procedures, equipment, and processes used to deliver medical care [9]. The range of such investments usually involves computer hardware and software that deal with the storage, retrieval, sharing, and use of health care information, data, and knowledge for communication and decision-making [10]. Although it is possible to identify stand-alone technologies under the H4.0 umbrella, they tend to be highly interrelated, generating the need to assess them in bundles. In addition, there is significant uncertainty regarding which technology will be the industry standard, adding an extra level of complexity to financial evaluations.

As the level of investment required to stay competitive with these new technologies is massive, the financial budgets of health institutions and countries are constantly stressed. For instance, data from the BRICS nations (ie, Brazil, Russia, India, China, and South Africa) indicate that their average health expenditure grew from 5.41% of their gross domestic product in 1995 to 6.94% in 2013 and is forecast to reach an average of 7.86% by 2025 [11]. Hence, there is an increasing need for massive and interconnected investments that will impose nontrivial challenges in determining their value, optimum level, and implementation sequence.

Several different theoretical lenses help to enlighten managers in their technological investments. The Health Technology Assessment International Policy Forum recently concluded that the assessment paradigms need to be more agile, helping health care systems to understand the potential of innovations and ensure that their potential value is realized [12]. However, although the literature has suffered from balkanization because multiple alternative approaches have grown significantly in recent years, hospitals rarely have, or use, a systematic decision process for H4.0 technology investments, accounting for all organizational objectives and using objective data [13,14].

This paper aims to address the current gap between the literature and practice by examining trends, challenges, and research opportunities in hospital investment valuations of H4.0 technologies. To achieve this goal, we opted to carry out a scoping review of the literature, which is appropriate for identifying and mapping critical concepts that underpin a specific research topic, especially in the absence of previous comprehensive studies [15,16]. More importantly, the scoping review approach is also suggested as an alternative to a systematic review when the literature is vast, sparse, and complex [17,18], which is the case of investments in H4.0 technologies [19].

The paper has been structured as follows. First, we motivate the study, present the protocol for the scoping review (ie, the research method section), and summarize the manuscript selection process. Second, we define the research questions, identify the relevant studies, and select the final list. Third, we present the main findings in a section devoted to the analysis of results, addressing the first two research questions. Fourth, we develop a framework that synthesizes the analysis and identifies promising research directions regarding the most crucial characteristics for evaluating investments in H4.0 technologies, addressing the third research question.

Hospital Investments, the Fourth Industrial Revolution, and Alternative Evaluation Approaches

The advent of Industry 4.0 technologies has significantly affected the global health care value chain. The recent integration of disruptive technologies derived from Industry 4.0 into health care systems aims at achieving virtualization to provide care in real time [20]. Health care institutions have incorporated cyberphysical systems, cloud computing, the Internet of Things, and big data, among others, into health care processes, services, equipment, material, and people. H4.0 technologies allowed the establishment of a smart system to monitor, track, and store patient records for ongoing care and analysis [21,22]. The combination of new technologies has expanded the scope of hospital activities. Economically, H4.0 technologies come with a value proposition of simultaneously improving efficiency and quality of care while reducing operating costs [23].

However, health care institutions need to carry out substantial investments to achieve the economic gains associated with H4.0 technologies. In 2014, US health care expenditure was US $3 trillion and is forecast to rise to US $5.1 trillion in 2023, outpacing the expected gross domestic product growth rate in the corresponding period [24,25]. These expenditures imply multiple investments that are not free of uncertainties because evaluating the impact on patient care is extremely difficult [26].

The unique characteristics of H4.0 technologies add a layer of evaluation complexity in an industry where assessing economic value is already challenging. For instance, studies on health technology assessment have primarily recognized that not every technological development results in net health gains [27]. The history of medicine and health includes many examples of technologies that did not produce the expected benefits or even proved harmful. At the same time, proving the effectiveness of technologies creates a continuous challenge for health systems because their application may require additional resources or compel health systems to choose from competing alternatives within the health system.

Studies have examined how health care organizations struggle to benefit from investments in H4.0 technologies [28,29]. Therefore, the dramatic increase in firms' technology investments in recent years has not necessarily resulted in a significant increase in productivity [30]. The complexity involved in understanding the economic impact of H4.0 technologies resulted in nontrivial challenges in determining the policy and practice implications associated with them [31].

Organizations contribute significant financial resources to developing and implementing H4.0 technologies, and the potential for a negative return on investments or total implementation failure is a worrisome possibility [32]. Assessing technological investments is of great interest to hospital managers when they seek to raise capital to expand services [33]. With the rapid growth of eHealth in developing countries, there is an urgent need for substantial evidence of its impact on justifying and guiding the investment of resources in such systems [26].

Studies evaluating H4.0 technology investments have taken different approaches. A wide array of manuscripts focus on cost reduction evaluation. For instance, a study by Galani and Rutten [34] reported that health care decision-makers base their adoption decisions on cost-effectiveness and cost-minimization analyses. The main limitation of this approach is the focus on just one aspect of the decision (cost), underemphasizing value considerations.

The real-options approach to decision-making has been useful in capturing and valuing uncertainty in many operating decisions that decision-makers face [35]. Its utility lies in the fact that real options are contingent on future discretionary investment. The magnitude, timing, and schedule of the investment outlay affect the value of firms' growth opportunities. Apart from correcting limitations from the cost perspective, the real-options approach increases the analytical effort that organizations need to carry out economic evaluations.

In addition, investing in H4.0 technologies requires multilateral stakeholder dialog and collaboration that address health needs and product conceptualization [12]. The nature of H4.0 technologies imposes challenges on how to assess the various aspects of technological value in the decision-making processes so that the assessment simultaneously accounts for the input of physicians, patients, and society [36].

Unsurprisingly, despite the expected benefits of H4.0 technologies and the interest from hospitals and policy makers in implementing them, the uptake and adoption of these technologies have not always been consistent within the health care practice, and adoption of these technologies has lagged [37]. There is a need to synthesize research activities and evidence to clarify the evaluation process of H4.0 technology investment in hospitals. Our scoping review explores this knowledge gap by mapping the extent and nature of the available literature and focusing on literature-based evidence that examined the integration of H4.0 technology investments into hospitals.

Methods

Overview

The scoping review design represents a methodology that allows the assessment of emerging evidence; therefore, it is the first step in research development [16]. It is a relatively new approach to evidence synthesis and differs from systematic reviews in its purpose and aims. The purpose of a scoping review is to provide an overview of the available research evidence without producing a summary answer to a discrete research question [38]. The methodology can help answer broad questions and gather and assess information before conducting a systematic review. It is suitable for achieving several objectives such as identifying the types of existing evidence in a given field, clarifying key concepts or definitions in the literature, surveying how research is conducted on a specific topic, identifying key characteristics related to a particular topic, and identifying knowledge gaps. Compared with systematic literature reviews and meta-analyses, a scoping review provides more flexibility and allows for diverse, relevant studies that use different methodologies [17,39,40]. Our research domain is adequate for performing a scoping review because studies regarding H4.0 technologies are multidisciplinary and relatively new.

To achieve our goal, we followed a standard scoping study procedure comprising five steps: (1) identify the research questions; (2) identify relevant studies; (3) select studies; (4) chart the data; and (5) collate, summarize, and report the results. In the following sections, we detail each stage and the outcomes of our study.

Identify the Research Questions

As with most systematic literature reviews, scoping reviews start with a primary research question to focus the inquiry [15,16], guiding researchers to build the search strategies [17]. Our broad initial research question was "How have health care institutions assessed their H4.0 technology investments?" However, given the multidisciplinary nature of the subject and the comprehensive sources of the reports, we narrowed the main research question into three more specific research questions:

Research question 1: What methodologies do health care institutions use for evaluating investments in H4.0 technologies?

Research question 2: What are the main challenges faced by health care institutions when evaluating investments in H4.0 technologies?

Research question 3: Which are the most important characteristics that the methodologies for evaluating investments in H4.0 technologies must have?

To answer these questions, we developed a rigorously structured and sufficiently documented method to provide robust evidence and arguments.

Identify Relevant Studies

A scoping review requires the identification of all relevant studies, regardless of the methodological design [16]. This step aims to find all available published and unpublished studies that address the research questions, operationalized through the search terms. As familiarity with the research topic is likely to increase as the review advances, we searched for relevant studies in two stages. In the first stage of identification, to include as many relevant studies as possible, we defined the set of keywords that best represented the scope of the study. In the second inclusion stage, we randomly selected a group of papers from each database and analyzed their keywords to determine the need to add more keywords to our inquiry. This two-stage process allowed us to address the search string's potential problem of being overly specific or entailing (partially) misleading buzzwords.

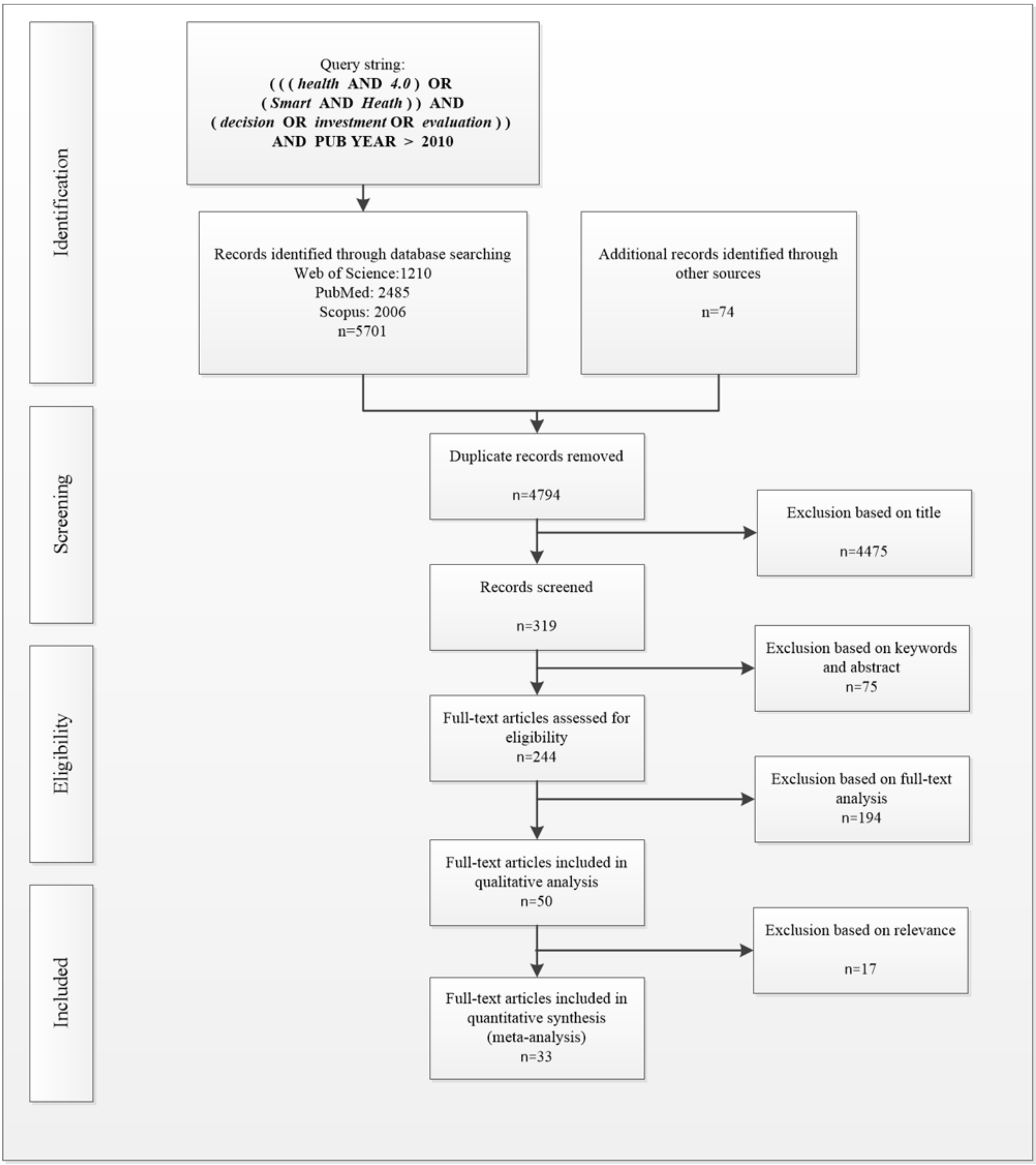
In the first stage, we defined the three research dimensions or keywords that best reflected our research questions: investment, health, and Industry 4.0. Subsequently, we combined an initial set of keywords using the AND and OR Boolean search operators (*investment* AND *health*, *health* AND *industry 4.0*, *investment* AND *industry 4.0*, *health technology assessment* AND *industry 4.0*) to retrieve publications that used them in the title, abstract, and keywords. The use of the AND operator in the search process significantly reduced misleading results, especially in the case of the *4.0* string. We searched for scientific articles in the following databases: Crossref, PubMed, Scopus, and Web of Science (which comprises biomedical literature from MEDLINE, life science journals, and web-based books).

As H4.0 derives from principles and technologies from Industry 4.0, whose concept was formally acknowledged in 2011 [41], we only considered publications after that year. Furthermore, in the widely referenced literature review by Liao et al [3], the authors indicated that, although the announcement of the Industry 4.0 concept traces back to April 2011, it began to attract attention only after it became one of the ten official projects within the *High-Tech Strategy 2020* action plan in March 2012. In fact, no study was identified before that date, supporting the choice of the cut-off year of 2011 for our scoping review.

We applied the query string to the indicated databases and retrieved a total of 5701 publications from these databases.

In the second stage, we randomly selected five articles from each database to compare their keywords with those from the research dimensions used in the first stage [42]. The objective was to take into account the fact that different taxonomies may be associated with a given subject, potentially compromising the search. From the comparisons, we identified the need to add the keyword *smart* to our inquiry. A new search, with this keyword included, generated 74 additional papers, giving us a total of 5775 publications scattered among the databases, as shown in Figure 1. We conducted both search stages from July 2000 to August 2020. Figure 1 charts the process of the identification of relevant studies and the final selection of the studies included in the review.

Figure 1. Selection of studies for the review.


Identification
Screening
Eligibility
Included
Query string:
(((health AND 4.0) OR
(Smart AND Heath)) AND
(decision OR investment OR evaluation))
AND PUB YEAR > 2010
Records identified through database searching
Web of Science:1210
PubMed: 2485
Scopus: 2006
n=5701
Additional records identified through other sources
n=74
Duplicate records removed
n=4794
Exclusion based on title
n=4475
Records screened
n=319
Exclusion based on keywords and abstract
n=75
Full-text articles assessed for eligibility
n=244
Exclusion based on full-text analysis
n=194
Full-text articles included in qualitative analysis
n=50
Exclusion based on relevance
n=17
Full-text articles included in quantitative synthesis (meta-analysis)
n=33

Selected Studies

The definition of different inclusion and exclusion criteria was post hoc because the researchers' familiarity with the studies increased. In the first exclusion process (screening), we considered only articles in English published in peer-reviewed journals. We removed duplicate publications from the portfolio, reducing the number of articles from the initial 5775 to 4794. In the next exclusion step, the paper titles were individually verified to determine their alignment with the research topic. This resulted in 4475 papers being deemed irrelevant to the research. The remaining 319 articles that passed the title screening were then checked for the alignment of keywords and abstracts with the research topic. A total of 75 articles were excluded, resulting in 244 articles being considered in the eligibility step.

The next step was to determine the eligibility of the papers. Best practice guidelines for conducting scoping reviews recommend that 2 separate reviewers carry out the literature search and sifting process. They must both agree before the study can be included. Therefore, we took special care to ensure interrater reliability, with at least two separate reviewers involved in the process.

We carried out the two separate review processes and performed a full-text analysis of the 244 articles to determine their eligibility. In all, 50 articles were identified by both reviewers as fully aligned with our research interests. We then evaluated the papers keeping in mind the criteria of relevance and methodological rigor. In this process, we added a third reviewer, and a majority vote determined the inclusion of a paper. By the end of this stage, 30 articles were considered appropriate for inclusion in the review. We also analyzed the references of these articles to identify relevant studies not yet included in the portfolio, but none were found. However, based on experts' recommendations (qualitative analysis), three articles were added to the portfolio, resulting in a final number of 33 studies in the publication portfolio, as displayed in Figure 1.

Results

Overview

We charted and interpreted critical data from the publication portfolio to establish the grounds for the subsequent analytical step [39]. We followed a descriptive-analytical method [17,43], providing a broader and meaningful view of all papers and collecting standard information from each study. Driven by our investigation's research questions, we organized the articles in a spreadsheet that included the following information: authors, year of publication, journal, aims, type of technology, application focus (eg, hospital processes or health treatments), valuation methods, decision-makers, users, challenges, and opportunities.

Table 1 shows a basic descriptive numerical summary of the publication count per year. Three main characteristics are noteworthy. First, as expected, studies on the financial evaluations of H4.0 technologies are recent. Second, there has been a slight increase in the number of publications in recent years (2018-2020). Finally, the number of included articles is relatively small (n=33), which may be due to the novelty of H4.0 technologies and the multidisciplinary nature of the investment evaluation requirements and its complexity. These findings reinforce the convenience of using a scoping review approach.

Table 1. Number of publications per year.

Year	Number of publications
2011	2
2012	1
2013	2
2014	4
2015	4
2016	3
2017	3
2018	6
2019	5
2020	3

Collate, Summarize, and Report Results

In this step, the results were collated, summarized, and reported based on a thematic framework such that a narrative account of the publication portfolio became available. Following the study by Levac et al [39], we carried out three complementary analyses to increase the consistency of this step. First, we performed a descriptive thematic analysis to collate and summarize the results. Second, based on the reported results, we developed a detailed analysis of the characteristics, contributions, and challenges of H4.0 technology evaluation tools. We report this analysis in the *Analysis of Results* section. In the *Classification Framework* section, we describe an emerging framework that synthesizes the empirical patterns of the analyzed papers. Finally, we discuss our findings' implications in a broader context, ensuring the scoping study methodology's legitimacy for both theory and practice [39]. In this discussion, we have also listed the research gaps and proposed research alternatives for future studies.

We now expand on the first step, providing detailed information on the characteristics of key publications. We conducted a word cloud analysis using the titles, keywords, and abstracts of papers in the portfolio. Multimedia Appendices 1-3 and Table 2 include the results, which provide initial evidence to answer the research questions. *Health* was the most frequent word, followed by *cost, cost-effectiveness, study, evaluation, care,patients,* and *data.*

The word cloud analysis anticipates the interdisciplinary nature of the manuscripts in the portfolio, allowing us to identify cost-effectiveness evaluation as the most recurrent. In addition, the incidence of the words *management*, *clinical*, and *patient* anticipates the need for health care institutions to incorporate a broad set of players in the investment decision process. We emphasize the absence of words such as quality, value, and bundle, which anticipate challenges and opportunities in current research on H4.0 technology investment analysis.

Table 3 reports the top 15 most frequent authors in the portfolio and the number of documents they authored, showing some of those who authored just one paper and the entire list of those who authored two or more. From the list of 146 authors, 1 participated in three studies, 2 participated in two studies, and the remaining 143 appeared in only one article. A large number of authors with small authoring prominence is typical of research topics about which knowledge is still incipient, such as H4.0 technology investment evaluation, reinforcing the convenience of adopting a scoping review as the methodological approach. We also observed a large average number of authors per publication (mean value of 4.33, SD 2.24), which is typical of publications in the medical field.

The journals' analysis also reinforced the topic's multidisciplinary nature. Table 4 reports the number of papers by category, and Table 5 reports the number of papers by journal. The Web of Science category *Health Care Sciences & Services* has the highest frequency of 16, followed by *Medical Informatics* (n=10), and *Pharmacology & Pharmacy* (n=2). The remaining categories displayed a frequency of 1 (6/9, 67% of the sample). Two journals published four manuscripts each: *Journal of Medical Internet Research* and *JMIR MHealth and UHealth*.

Table 2. Most frequent words in titles and abstracts.

Section and word	Count
Title	
Health	12
Evaluation	10
Cost	8
Cost-effectiveness	7
Effectiveness	7
Study	7
Based	5
Decision	5
Economic	5
Management	5
Keyword	
Health	31
Results	26
Methods	24
Study	23
Analysis	20
Based	19
Care	19
Data	17
Background	16
Cost	16

Table 3. Top 15 authors and frequency of appearance in publications (alphabetically ordered).

Author	Frequency
Wernz, Christian	3
Trajkovik, Vladimir	2
Zhang, Hui	2
Abrams, Keith R	1
A'Court, Christine	1
Adem, Abdu	1
Aloui, Saber	1
Augusto, Vincent	1
Babar, Zaheer Ud Din	1
Baio, Gianluca	1
Ball, Daniel R	1
Belani, Hrvoje	1
Berta, Whitney	1
Bertranou, Evelina	1
Beutner, Eric	1

Table 4. Frequency of manuscripts stratified according to Web of Science category.

Web of Science category	Frequency
Health care sciences and services	16
Medical informatics	10
Pharmacology and pharmacy	2
General and internal	1
Information systems	1
Computer sciences	1
Medicine	1
Multidisciplinary sciences	1
Operations research and management	1

Table 5. Frequency of manuscripts by journal.

Journal	Frequency
Journal of Medical Internet Research	5
JMIR mHealth and uHealth	4
The Oncologist	2
Applied Health Economics and Health Policy	1
Biomedical Instrumentation and Technology	1
BMC Health Services Research	1
BMJ Open	1
Clinical Therapeutics	1
Frontiers in Pharmacology	1
Global Health Science and Practice	1
Health Affairs	1
Health Care Management Science	1
Health Economics Review	1
Implementation Science	1
Industrial Management and Data Systems	1
Information Systems and e-Business Management	1
International Journal of Medical Informatics	1
International Journal of Operations and Production Management	1
Journal of Multidisciplinary Healthcare	1
Journal of Operations Management	1
Journal of the Canadian Academy of Child and Adolescent Psychiatry	1
Journal of the Royal Statistical Society. Series A: Statistics in Society	1
PLOS ONE	1
Proceedings of the 51st Hawaii International Conference on System Sciences	1
Value in Health	1

Analysis of Results

Tables 6 and 7 summarizes the papers listed in the rows by year of publication and their different content characteristics. We started by reporting the type of technology analyzed, grouping them according to their role within the health care organization. The study by Aceto et al [5] proposed four interrelated subsets: (1) communication, (2) sensing, (3) processing, and (4) actuation. Communication involves different interactions and disseminating health-related information, supporting patient-professional relationships, and providing collaborative care. Related H4.0 technologies provide support to increase accessibility, exchange, and sharing of information. Sensing refers to acquiring information about a patient, equipment, material, or process without necessarily making physical contact with them. Processing refers to technologies that may change or process the acquired data, producing actual information in any manner detectable by an observer. Finally, actuation refers to technologies responsible for moving and controlling a system, mechanism (electronic or mechanical), or software based on the information and signals received.

Table 6. Classification of contents in the portfolio of papers.

Study and year	Non Healthcare 4.0	Healthcare 4.0		Technologies		Valuation methods					
		Sensing and communication	Processing and actuation	Stand alone	Bundle or portfolio	Deterministic		Uncertainty no option		Uncertainty option analysis	
						Cost	Value	Cost	Value	Cost	Value
Dreyfuss and Roberts [44] (2011)			✓[a]	✓							✓
Grutters et al [45] (2011)			✓	✓							✓
Marsh et al [46] (2012)	✓			✓							✓
Favato et al [47] (2013)				✓							✓
Drummond et al [48] (2013)	✓			✓		✓					
Pertile et al [49] (2013)	✓							✓			
Boydell et al [50] (2014)		✓		✓							
Kvedar et al [51] (2014)		✓	✓	✓	✓						
Wernz et al [14] (2014)		✓	✓	✓							✓
Atwood et al [52] (2015)		✓	✓	✓							✓
Wernz et al [13] (2015)		✓	✓	✓							✓
Gobbi and Hsuan [53] (2015)		✓		✓				✓			
Merlo et al [54] (2015)								✓			
Sharma et al [55] (2016)		✓	✓		✓						
Matthew-Maich et al [56] (2016)		✓	✓	✓							
de Grood et al [37] (2016)		✓		✓							
Lavallee et al [57] (2017)		✓			✓						
Kim and Lee [58] (2017)		✓	✓	✓							
Rejeb et al [59] (2017)	✓				✓			✓			
Greenhalgh et al [60] (2017)		✓	✓	✓		✓					
Long et al [61] (2018)		✓			✓						
Adjekum et al [62] (2018)		✓		✓							

Study and year	Non Healthcare 4.0	Healthcare 4.0		Technologies		Valuation methods					
		Sensing and communication	Processing and actuation	Stand alone	Bundle or portfolio	Deterministic		Uncertainty no option		Uncertainty option analysis	
						Cost	Value	Cost	Value	Cost	Value
Winters et al [63] (2018)		✓		✓							
Baines et al [64] (2018)		✓	✓	✓							
Taj et al [65] (2019)		✓	✓	✓							
Dogba et al [66] (2019)		✓		✓							
Loncar-Turukalo et al [67] (2019)		✓	✓	✓	✓						
Shahid et al [68] (2019)		✓	✓	✓							
Wüller et al [69] (2019)		✓	✓	✓							
Chouvarda et al [70] (2019)		✓		✓				✓			
Hasselgren et al [71] (2020)		✓									
Peng et al [72] (2020)		✓	✓	✓							
Ismail et al [73] (2020)		✓	✓								

[a] Present in study.

Table 7. Classification of some contents in the portfolio of papers.

Study and year	Decision maker			User		
	Medical	Administrative	Patient	Medical	Administrative	Patient
Dreyfuss and Roberts [44] (2011)		✓[a]			✓	
Grutters et al [45] (2011)		✓			✓	
Marsh et al [46] (2012)		✓			✓	
Favato et al [47] (2013)	✓	✓				
Drummond et al [48] (2013)	✓			✓		
Pertile et al [49] (2013)		✓			✓	
Boydell et al [50] (2014)	✓		✓			✓
Kvedar et al [51] (2014)	✓		✓			✓
Wernz et al [14] (2014)	✓	✓			✓	
Atwood et al [52] (2015)	✓	✓			✓	
Wernz et al [13] (2015)		✓			✓	
Gobbi and Hsuan [53] (2015)		✓			✓	
Merlo et al [54] (2015)	✓	✓	✓	✓	✓	
Sharma et al [55] (2016)	✓			✓		
Matthew-Maich et al [56] (2016)	✓	✓		✓		
de Grood et al [37] (2016)	✓	✓		✓	✓	
Lavallee et al [57] (2017)	✓			✓		
Kim and Lee [58] (2017)	✓		✓			✓
Rejeb et al [59] (2017)	✓			✓		
Greenhalgh et al [60] (2017)			✓			✓
Long et al [61] (2018)	✓			✓		
Adjekum et al [62] (2018)	✓			✓		
Winters et al [63] (2018)	✓		✓	✓		✓
Baines et al [64] (2018)	✓		✓	✓		✓
Taj et al [65] (2019)	✓			✓		
Dogba et al [66] (2019)	✓		✓	✓		✓
Loncar-Turukalo et al [67] (2019)	✓			✓		
Shahid et al [68] (2019)	✓			✓		
Wüller et al [69] (2019)	✓			✓		
Chouvarda et al [70] (2019)	✓			✓		
Hasselgren et al [71] (2020)	✓	✓		✓	✓	
Peng et al [72] (2020)	✓			✓		✓
Ismail et al [73] (2020)	✓	✓		✓	✓	

[a] Present in study.

There may be overlaps among the technology subsets. Following the classification in the study by Tortorella et al [1], we further grouped H4.0 technologies into two bundles according to their role within the hospital: sensing-communication (reported under the column labeled *sensingand communication*) and processing-actuation (reported under the column labeled *processing andactuation*; Table 6). Consistent with previous studies' reports on the incidence of technological applications (eg, the study by Tortorella et al [1]), the number of articles evaluating sensing-communication is significantly greater than those analyzing processing-actuation. In addition, and somewhat paradoxically, given the nature of H4.0 technologies, most studies focus on just one technology, with only six manuscripts addressing bundles of technologies.

Regarding the thematic analysis (data not included in Table 6 because of space limitations), we observed two groups of studies on H4.0 technology evaluation in health care organizations: those related to health treatments and those related to hospitals' supporting and administrative processes. Articles in the former group were relatively more frequent than those in the latter group.

The evaluation of different technologies contributes to health improvement in various ways. H4.0 technologies contribute to reductions in diseases such as cancer [44,59] and allow for better connectedness that manages individual and community health holistically by leveraging various technologies [70]. Connectedness can also incorporate telehealth and integrated care services, covering the entire spectrum of health-related services that address healthy individuals and patients with chronic conditions [70]. In addition, neural networks improve decision-making, improving care delivery at a reduced cost [68].

The aforementioned analysis allowed us to describe the types of technology and the health improvement sought from their use. Next, we addressed the first research question. For this purpose, we surveyed the methodologies that health care institutions reportedly use for evaluating investments in H4.0 technologies.

With regard to the different methodologies for evaluating investments in H4.0 technologies, of the 33 papers analyzed, only 14 (42%) presented valuation methods, whereas 7 (21%) focused on cost valuation methods and 7 (21%) focused on value methods. Regarding forms of considering uncertainty in the analysis, of the 33 papers analyzed, 2 (6%) used deterministic techniques that disregarded uncertainties, 5 (15%) accounted for uncertainty but did not use real options, and 7 (21%) accounted for uncertainty using a real-options approach.

As we can observe, studies that consider the cost implications of investing in H4.0 technologies focus on economic analysis, adopting a cost-effectiveness and cost-minimization perspective. These studies were complemented by the application of a Bayesian sequential economic evaluation model for health technologies in which an investigator has flexibility over the timing of a decision to stop carrying out research and conclude that one technology is preferred over another on cost-effectiveness grounds [49]. A total of five manuscripts took a real-options perspective that incorporates value considerations but refers to past work, mainly published at the beginning of the time window of analysis.

The portfolio of 33 studies lists three types of decision-makers, who may be consulted individually or in groups: doctors, administrative staff, and patients. Doctors appear in 82% (27/33) of the studies; 39% (13/33) incorporate the administrative perspective; and 21% (7/33) contain the patient perspective. Although there is a dominance of expert opinion based on medical advice, the variety of decision-makers is a positive result that further claims support for a multidisciplinary analysis that incorporates the different types of users affected when evaluating investments in H4.0 technologies. Users of the information derived from the evaluations are also doctors, administrative staff, and patients; however, administrative users are predominant because they are direct users of the economic information.

A relevant aspect of the 33 studies analyzed in the portfolio is that 48% (16/33) of them present results of scoping or systematic literature reviews and meta-analyses (1 meta-analysis present among the 16 papers). However, they focused on the medical convenience of H4.0 technology investments, not on exploring specific economic evaluation tools, and mainly assessed a particular technology (eg, physicians' adoption of eHealth technology or smart device apps for older adults).

We were able to consolidate several relevant propositions for the economic evaluation of H4.0 technologies. A fundamental contribution of our review is the identification of the main antecedents of hospital investment decisions in technology, such as the health care system, the socioeconomic and cultural context, and its mission [13,14]. Regarding the health care system, the findings emphasize the role of health insurance coverage, financing methods, reimbursement methods for hospitals, methods used to make payments to physicians, and hospital ownership as antecedents of H4.0 technology investments. The existence of these antecedents anticipates the challenges of investment evaluations [14,37,44-74].

The appropriate deployment of medical technology should help to contribute to the quality of health care delivered, improve access to information, and contain costs [52]. Among the most promising evaluation alternatives is the framework in the study by Greenhalgh et al [60] to assist implementation teams in identifying, understanding, and addressing the interacting challenges to achieving sustained adoption, local scale-up, distant spread, and long-term sustainability of their technology investments in hospitals. Complementing this analysis is the call for applying a simple, multiattribute rate technique in the valuation process, as proposed in the study by Wernz and Zhang [13].

We identified four main challenges faced by health care institutions when evaluating investments in H4.0 technologies. First, H4.0 technologies should be analyzed as a bundle of technologies rather than individual solutions. As proposed in the study by Aceto et al [5], there are four overlapping groups of technologies based on their roles and applicability within the hospital. In our portfolio, of the 33 papers, only 6 (18.1%) analyzed H4.0 technologies as a bundle. Second, as mentioned earlier, there is a research gap in valuation methodologies for H4.0 technologies, especially in the realm of real-options analysis. Third, regarding who makes the decision to acquire the technology (medical personnel, administrative staff, or patient), 82% (27/33) of studies focused on the medical personnel as the main decision-makers. In contrast, only 24% (8/33) focused on patients, and a single paper integrated the 3 actors in the process [54]. Fourth, regarding the user of the technologies, 67% (22/33) of studies focused on medical personnel, whereas 24% (8/33) indicated that the main user was the patient.

Real-options strategies offer a transparent method for weighing the costs and benefits of adopting and further researching new and expensive technologies [44,45]. Such valuation methodologies incorporate the value of future new information in the current analyses. The articles in the portfolio report real-options applications in proton therapy adoption analysis

[44,45] and help formulate better human papillomavirus vaccination strategies [47]. Surprisingly, none of the articles using real-options analysis incorporated uncertainty correlations among the bundles of technologies. This is a critical shortcoming given the antecedents that report the importance of taking broader portfolio considerations when evaluating related and uncertain investments in areas such as biotechnology research and development [75].

Health care managers often make purchasing decisions without adequately assessing the resource demands, up-front costs (including integration costs), workflow impact, reimbursement potential, and other factors needed to fully understand the value added by new medical technology [52]. Consequently, health care authorities may risk failing to conduct thorough due diligence before purchasing medical technology. Under these circumstances, organizations might add unnecessary costs to their budget without adding significant clinical or operational value.

Selecting new medical technology for a health care organization can be a daunting task. It is crucial to implement a systematic approach for evaluating the latest medical technology, starting with a clearly articulated need for the technology. If organization authorities are unwilling to assess and redesign processes to fully use the new medical technology, investment withholding may be the most suitable course of action. Moreover, there is a risk of bias in purchasing the latest technology simply because it is available [52]. Overall, health care organizations rarely assess a systematic decision process that considers all organizational objectives and analyzes and integrates comprehensive data [52].

Providing universal access to innovative, high-cost technologies has led to tensions in today's health care systems. The stress becomes particularly evident in the context of scarce resources, where the risk of taking contentious coverage decisions increases rapidly. If health care institutions intend to maintain sustainable access to H4.0 technologies in the future, new approaches are needed to reconcile these different perspectives [48]. Overall, although policy makers request rapid and at-scale technology implementation, the reality is that when dealing with the multiple complexities of health and care, it is challenging to go beyond small-scale demonstration projects [52]. To address the need for new approaches, we propose in the next section a framework for the evaluation of H4.0 technologies in hospitals.

Classification Framework

Scientific research presents frameworks because managers use them to support their analysis and provide validity to the decision-making process [76]. We developed an emerging framework from the study we conducted on the research on hospital evaluations of H4.0 technologies.

Frameworks have multiple advantages. They decrease the number of uncertainties when addressing a new phenomenon, as is the case with H4.0 technologies. Frameworks can support the selection of investment strategies. In addition, frameworks can depict features of various phenomena [77], compare and guide numerous organizational practices [78], support the execution of tasks [79], and refute or confirm a particular management approach [80].

When developing the framework, it is fundamental to determine the rationale that validates the theoretical process. Given the scoping review's multidisciplinary and integrative nature, we have chosen a process of abstraction, that is, we obtain higher-order themes from lower-order elements [81]. Therefore, we follow the most common abstraction process, in which lower-order themes are a function of the findings of individual studies, and higher-order structures link and organize the lower-order themes [81]. Such a method should result in the advancement of knowledge rather than a simple overview or description of a research area [82], that is, it should not be descriptive or historical but should preferably generate a new conceptual framework. In addition, we checked the reliability of higher-order themes using a focus group of experts. It is worth noting that the higher-order themes respond to taxonomy and not from a typological process [82,83].

Figure 2 presents the proposed classification framework. It focuses on the most fundamental tensions that organizations face when analyzing H4.0 technology investments and reflects the most prominent features of our publication portfolio. We categorized the type of technology analyzed based on its focus, sensing-communication or processing-actuation, following the classification in the study by Tortorella et al [1]. In this process, we classify lower-order themes into higher-order classification. We describe the number of technologies evaluated, depending on whether the analysis refers to stand-alone technologies or bundles. We also report the evaluation method, stating whether it is based only on cost or also takes into account value considerations. We considered whether the analysis does not incorporate flexibility in the valuation process or explicitly incorporates it using a real-options approach. We also examine the portfolio of manuscripts regarding the variety of decision-makers included and the type of technology users. For all these cases, we propose higher-order themes for the portfolio of manuscripts.

The framework not only helps to classify a particular research paper but also has utility for practice. It may allow hospital authorities to understand what type of organizational process they have in place to analyze investment decisions in H4.0 technologies. In addition, it helps to anticipate the complexity of the task. When reflecting on the most critical tensions that hospitals face, the structure would allow authorities to detect the underlying leadership and change-management challenges.

When categorizing the portfolio of manuscripts using the proposed framework, we identified a significant concentration of studies on the left side. It seems reasonable to observe such an unbalanced distribution, given the developing nature of H4.0 technologies. However, it also signals an essential shortcoming of the current studies, directing further research propositions. There is a risk that hospitals might have been making decisions by following isomorphic behavior [84], which is not necessarily the best rational approach. Research concentration might reflect herding behavior in which hospitals imitate one another instead of following a robust, innovative path.

We further analyzed the framework and developed a research opportunity map, displayed in Figure 3, focusing on two dimensions of the framework: the complexity of the analysis and the number of technologies considered.

From the research map, it is possible to indicate that there is a research opportunity related to analyzing bundles of technologies with complex relationships that incorporates uncertainty correlations. It is important to emphasize that complex relationships do not necessarily imply more complex analyses. The challenge is to integrate a higher level of complexity with straightforward analytical tools. We will return to this point at the end of the following section.

Figure 2. Classification framework.

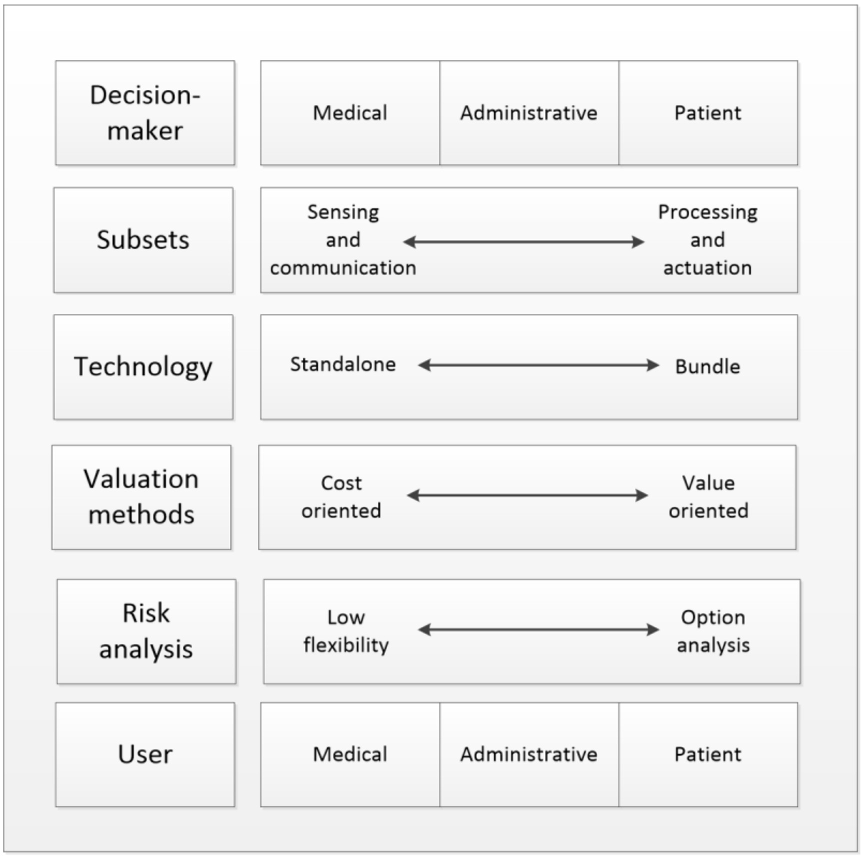
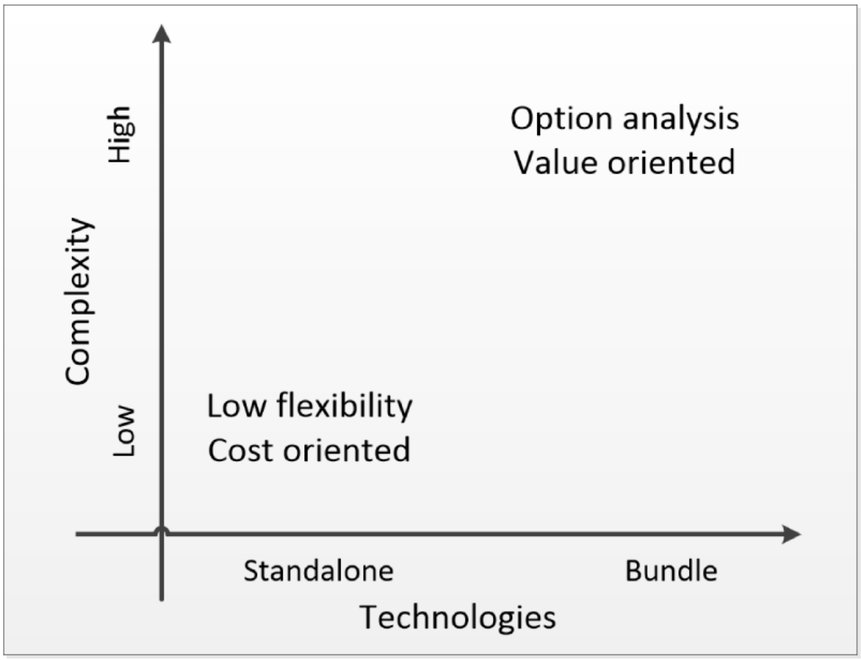

Decision-maker
Medical
Administrative
Patient
Subsets
Sensing and communication
Processing and actuation
Technology
Standalone
Bundle
Valuation methods
Cost oriented
Value oriented
Risk analysis
Low flexibility
Option analysis
User
Medical
Administrative
Patient

Figure 3. Research opportunity map.


Complexity
High
Low
Option analysis
Value oriented
Low flexibility
Cost oriented
Standalone
Bundle
Technologies

Discussion

Principal Findings

This section addresses the third research question, that is, we identify the crucial characteristics that methodologies for evaluating investments in H4.0 technologies should have. These characteristics also represent research gaps that should be addressed in future research.

Overall, we anticipate that the evaluation of H4.0 technologies presents challenges and opportunities that are similar to those related to general technology investments, although the evaluation is more complex because of the nature of H4.0 technologies. We also observe that the existing research does not entirely succeed in helping hospitals in the investment decision-making process, leading to promising research opportunities.

Insufficient Economic Valuation of H4.0 Technology Investments

Decades of research on health technology assessment have resulted in a framework that includes economic evaluation as a fundamental pillar. However, studies that rigorously integrate this economic perspective are still scarce. Advancements focus more on cost-effectiveness than economic value at the public policy level rather than at the hospital level. In addition, studies have confirmed that health care institutions rarely apply a systematic analysis that considers all organizational objectives and integrates comprehensive data [13]. For instance, the relatively low commercial externality valuation is one of the shortcomings of economic analysis.

Although some older studies on real options include cost and value considerations, more recent propositions tend to overfocus on cost analysis, imposing a challenging bias on investment decisions. Among propositions that incorporate value, the net present value analysis is the most frequently used, often resulting in a suboptimal decision because it does not consider the value of future options and managerial flexibility [13]. Usually, simple cost-benefit analysis and subjective assessment replace sophisticated analytical methods and objective data at the risk of not investing in more expensive technologies with higher health impacts because of their investment requirements. The development of real-options approaches that include value considerations targeted at evaluating investments in H4.0 technologies is a promising research opportunity that should resonate positively among practitioners.

Explicit Assessment of Technological Interrelationships

The literature provides evidence that for maximizing the return on H4.0 technology investments, hospitals should consider them in bundles. Studies have proposed distinct bundles (or groupings) of H4.0 technologies. Sharma et al [55] categorized technologies into three bundles according to the extent of patient-centered integration and caregiver interaction. The study by Aceto et al [5] conceptually proposed four overlapping groups of technologies based on their roles and applicability within the hospital. The study by Gastaldi and Corso [85] proposed another categorization of H4.0 technologies, dividing them into four macroareas, further subdivided into 14 solutions provided by each technology. Finally, the study by Alrige and Chatterjee [86] suggested a taxonomy to classify wearable technologies in health care systems according to three major dimensions: application, form, and functionality.

Although the literature still lacks consensus on the correct taxonomy of bundles of H4.0 technologies and how to combine them to act synergistically, it is clear that the valuation should incorporate the bundling of technologies. Therefore, researchers and institutions need to assess portfolio effects explicitly [75]. The literature on real options includes several studies that explicitly address portfolio considerations [75,87-89], providing a potential area of extension to H4.0 technology investments. In analyzing hospital investments, research incorporating portfolio considerations is scarce (eg, the study by Wernz and Zhang [13]) and does not include real-options valuations. It is fundamental to understand whether investing in technology bundles creates super- and subadditivity [75], altering the net economic contribution of different alternatives and eventually changing the suggested priorities.

We detected recent efforts to provide an accessible and usable framework that would enable multiple objectives, mainly developed by authors seeking to design, develop, implement, scale-up, spread, and sustain technology-supported health or social care programs to identify and help address the critical challenges in different domains and the interactions among them [51,60,67]. However, the developments only start to address the shortcomings identified in our scoping review, opening opportunities for future research.

Incorporate Fundamental Uncertainties

H4.0 technologies enhance efficiency and quality in health care systems. However, fundamental uncertainties exist in the definition of industry standards for many of these technologies, creating uncertainty when evaluating investments. Factors that add additional complexities to technological advancements relate to uncertainty regarding patient demands and competition [13].

To reduce the risk of investing in a technology that ends up being crowded out and not adopted as the standard, hospitals have several alternatives; further research is needed to explore their viability. Surprisingly, the discussion about standards is scarce in the economic evaluation of H4.0 technologies, with the main focus continuing to be on their efficacy.

Integrating Administrative, Medical, and Patient Perspectives in the Evaluation Process

The fourth research opportunity relates to integrating medical, patient, and administrative considerations in the valuation process. We have already stressed that the interrelationships among technology bundles incorporate nontrivial challenges. In addition, institutions should consider the risk of investing in technologies that fail to establish the industry standard. The final layer should adequately balance medical benefits with economic costs. It is still unclear how to achieve such reconciliation [48]. The central problem concerns the resolution of the economic logic versus medical logic debate. On the one hand, doctors favor technologies with the most promising medical effects, regardless of uncertainty and varying requirements of investment and cost. On the other hand, the administrative staff need to ensure the hospital's economic viability. Amid high levels of uncertainty, the amount of investment and the operating costs, that is, the economic logic, might contradict the medical logic. Research is needed to explore the most suitable ways in which hospitals can coordinate both perspectives.

In integrating the different perspectives into the valuation process, hospitals need to include those of the patient for at least three reasons [48]. First, a comprehensive assessment should consider patients' views on the satisfaction and acceptability of health technologies. Second, with chronic forms of disease and disability, patients and their families play a more active role in health care decisions. Patients' lifestyles and behaviors may dramatically influence long-term prognoses of chronic conditions. Third, the involvement of patients increases transparency and openness in public policy [48]. We acknowledge that incorporating the patient's view in the investment decision analysis adds a layer of complexity to a process that is already difficult to manage. However, any valuation analysis that considers costs and value without including the patient perspective will be incomplete.

The integration of different perspectives provides an opportunity to cross-fertilize research on H4.0 technology investments with adaptive leadership tools [90]. Alternatively, the incorporation of H4.0 technologies equals establishing a dynamic organizational capability that demands from employees the ability to leverage interpersonal relationships conducive to productive dialog [91].

Remain Manageable in the Decision-Making Process

Previous studies describe hospitals' investment decisions as ad hoc, informal, political, without sufficient data analysis, and not aligned with the institutions' mission and strategy [13]. We argue in favor of assessments that explicitly consider technological interrelationships, incorporate fundamental uncertainties, and integrate administrative and medical insights. However, our argument comes with an essential caveat: analytical methods should avoid introducing complex evaluation tools that hamper the hospital's decision-making process.

At first, such a requirement seems to be challenging. We have suggested incorporating bundles of technologies, mapping multiple uncertainties, considering value implications and not exclusively cost aspects, and including different stakeholders' perspectives. A priori, these requirements go against the simplification of the decision-making process. However, it might be possible to solve this tension by articulating the valuation process in different stages. We envision a lean financial valuation that combines these competing demands without drastically complicating the decision process.

The lean financial valuation of H4.0 technology investments involves simplifying, streamlining, and harmonizing essential valuation processes to create a leaner, more efficient valuation operation. The current research opportunity relates to developing lean organizations that incorporate valuation tools that simultaneously address challenges such as complex uncertainty relationships and bundle effects into organizational structures that adjust to lean principles.

Limitations and Final Remarks

This study examined how hospitals approach investment decisions in H4.0 technologies by using a scoping review of the existing literature. We performed a search for journal articles in four databases and screened relevant contributions to consolidate a publication portfolio on the topic, following predefined criteria. The results of the scoping review were explored using the following stepwise approach:

1. A descriptive numerical summary and thematic analysis.
2. Identification of trends and challenges in H4.0 technology investment evaluation.
3. Proposal of a classification framework for H4.0 technology investment evaluation.
4. Identification of research opportunities and proposals for future research directions from a hospital investment management point of view.

Despite the topic's recency, we observed that research in H4.0 technologies expands interdisciplinarily with a diversified set of applications and functionalities. In terms of the economic evaluation, studies on H4.0 technologies tend to overfocus on cost considerations and underemphasize cost-value relationships. Studies that consider both sides of the economic valuation (ie, value and cost) use real-options analysis and tend to be older in the sample of studies analyzed. Although the impacts of H4.0 technology adoption substantially increase when hospitals adopt technologies in bundles, research mainly focuses on the analysis of single technologies. Finally, recent studies have called for the integration of different actors in the decision process by developing a comprehensive, consistent, and data-driven framework for evaluating hospitals' investment decisions. We have proposed a framework that serves as a starting point.

Our study includes some noteworthy limitations, mostly related to its nature and methodological choices. As Industry 4.0 was formally acknowledged in 2011 and H4.0 is a concept derived from it, our scoping study only encompassed studies after that year. However, it is worth mentioning the existence of initiatives aimed at valuing Industry 4.0 technologies in health care systems not characterized as such and dating earlier than 2011, which is a limitation of our research. Nevertheless, because studies before 2011 were scarce and scattered, and the number of publications on the topic has significantly increased in the past few years, we believe that our choice of the search period returned all relevant studies on H4.0 technologies.

A second limitation is that we focused our literature analysis and discussion on H4.0 technology evaluation within hospitals. However, the concept of health care has expanded beyond the limits of health care organizations (ie, hospitals and clinics). In fact, with the advent of *smart cities*, complementary aspects of health care have been integrated because of the increased level of interconnectivity and data acquisition, allowing health care services to be demanded remotely. Our study did not analyze these aspects and exclusively considered hospitals the units of analysis.

Third, it is worth emphasizing that we combined insights from two perspectives to develop the proposed framework: the state of the practice at hospitals and the state of the art in the literature. However, our main focus was on research, and we did not include a specific survey of empirical studies mapping hospital tools. This is simultaneously a limitation of our investigation and a research opportunity.

Finally, identifying trends, challenges, and theoretical gaps through this scoping review allowed us to develop a framework. However, we acknowledge that this is the first step toward the proposal of an in-depth framework. Future studies could use the theoretical consolidation of the studies in our paper as a conceptual baseline for developing such a detailed H4.0 technology evaluation framework. We hope that our classification framework will act as a solid starting point for future developments in evaluating H4.0 technology investments.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Word cloud analysis using titles as input.
[PNG File , 93 KB - jmir_v23i8e27571_app1.png]

Multimedia Appendix 2
Word cloud analysis using keywords as input.
[PNG File , 91 KB - jmir_v23i8e27571_app2.png]

Multimedia Appendix 3
Word cloud analysis using abstracts as input.
[PNG File , 95 KB - jmir_v23i8e27571_app3.png]

References

1. Tortorella G, Fogliatto F, Vergara A, Vassolo R, Sawhney R. Healthcare 4.0: trends, challenges and research directions. Prod Plan Control 2019 Dec 17;31(15):1245-1260. [doi: 10.1080/09537287.2019.1702226]
2. Dehe B, Bamford D. Quality function deployment and operational design decisions – a healthcare infrastructure development case study. Prod Plan Control 2017 Jul 13;28(14):1177-1192. [doi: 10.1080/09537287.2017.1350767]
3. Liao Y, Deschamps F, Loures ED, Ramos LF. Past, present and future of Industry 4.0 - a systematic literature review and research agenda proposal. Int J Prod Res 2017 Mar 28;55(12):3609-3629. [doi: 10.1080/00207543.2017.1308576]
4. Xu LD, Xu EL, Li L. Industry 4.0: state of the art and future trends. Int J Prod Res 2018 Mar 09;56(8):2941-2962. [doi: 10.1080/00207543.2018.1444806]
5. Aceto G, Persico V, Pescapé A. The role of Information and Communication Technologies in healthcare: taxonomies, perspectives, and challenges. J Netw Comput Appl 2018 Apr;107:125-154. [doi: 10.1016/j.jnca.2018.02.008]
6. Pramanik M, Lau R, Demirkan H, Azad M. Smart health: Big data enabled health paradigm within smart cities. Expert Syst Appl 2017 Nov;87:370-383. [doi: 10.1016/j.eswa.2017.06.027]
7. Byrne CM, Mercincavage LM, Pan EC, Vincent AG, Johnston DS, Middleton B. The value from investments in health information technology at the U.S. Department of Veterans Affairs. Health Aff (Millwood) 2010 Apr;29(4):629-638. [doi: 10.1377/hlthaff.2010.0119] [Medline: 20368592]
8. Wang Y, Kung L, Wang W, Cegielski C. An integrated big data analytics-enabled transformation model: application to health care. Inf Manag 2018 Jan;55(1):64-79. [doi: 10.1016/j.im.2017.04.001]
9. May C, Finch T, Mair F, Ballini L, Dowrick C, Eccles M, et al. Understanding the implementation of complex interventions in health care: the normalization process model. BMC Health Serv Res 2007 Sep 19;7:148 [FREE Full text] [doi: 10.1186/1472-6963-7-148] [Medline: 17880693]
10. Angst CM, Block ES, D'Arcy J, Kelley K. When do IT security investments matter? Accounting for the influence of institutional factors in the context of healthcare data breaches. MIS Q 2017 Mar 3;41(3):893-916. [doi: 10.25300/misq/2017/41.3.10]
11. Jakovljevic M, Potapchik E, Popovich L, Barik D, Getzen TE. Evolving health expenditure landscape of the BRICS nations and projections to 2025. Health Econ 2017 Jul 29;26(7):844-852. [doi: 10.1002/hec.3406] [Medline: 27683202]
12. Husereau D, Henshall C, Sampietro-Colom L, Thomas S. Changing health technology assessment paradigms? Int J Technol Assess Health Care 2016 Oct 21;32(4):191-199. [doi: 10.1017/s0266462316000386]
13. Wernz C, Zhang H. Medical technology investment decision-making at U.S. Hospitals: a comparative case study of four organizations. In: Proceedings of the Hawaii International Conference on System Sciences 2018. 2018 Presented at: Hawaii International Conference on System Sciences; 2018; Waikoloa Village, Hawaii, United States. [doi: 10.24251/hicss.2018.388]

14. Wernz C, Zhang H, Phusavat K. International study of technology investment decisions at hospitals. Ind Manag Data Syst 2014;114(4):568-582. [doi: 10.1108/imds-10-2013-0422]
15. Mays N, Roberts E, Popay J. Synthesising research evidence. In: Fulop N, Allen P, Clarke A, Black N, editors. Methods for Studying the Delivery and Organisation of Health Services. London, United Kingdom: Routledge; 2001.
16. Peterson J, Pearce P, Ferguson L, Langford C. Understanding scoping reviews: definition, purpose, and process. J Am Assoc Nurse Pract 2017 Jan;29(1):12-16. [doi: 10.1002/2327-6924.12380] [Medline: 27245885]
17. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
18. Grimshaw J. A guide to knowledge synthesis: a knowledge synthesis chapter. Canadian Institutes of Health Research. 2010. URL: https://cihr-irsc.gc.ca/e/41382.html [accessed 2021-07-30]
19. Pan J, Ding S, Wu D, Yang S, Yang J. Exploring behavioural intentions toward smart healthcare services among medical practitioners: a technology transfer perspective. Int J Prod Res 2018 Nov 29;57(18):5801-5820. [doi: 10.1080/00207543.2018.1550272]
20. Sannino G, De Falco I, Pietro GD. A Continuous Noninvasive Arterial Pressure (CNAP) approach for Health 4.0 systems. IEEE Trans Ind Inf 2019 Jan;15(1):498-506. [doi: 10.1109/tii.2018.2832081]
21. Elhoseny M, Abdelaziz A, Salama AS, Riad A, Muhammad K, Sangaiah AK. A hybrid model of Internet of Things and cloud computing to manage big data in health services applications. Futur Gener Comput Syst 2018 Sep;86:1383-1394. [doi: 10.1016/j.future.2018.03.005]
22. Wu F, Li X, Xu L, Kumari S, Sangaiah AK. A novel mutual authentication scheme with formal proof for smart healthcare systems under global mobility networks notion. Comput Electric Eng 2018 May;68:107-118. [doi: 10.1016/j.compeleceng.2018.03.030]
23. Matopoulos A, Michailidou L. Implementing collaborative practices in the healthcare supply chain: insights into hospital-vendor operations. Int J Logist Syst Manag 2013;15(2/3):288. [doi: 10.1504/ijlsm.2013.053773]
24. Atasoy H, Chen P, Ganju K. The spillover effects of health IT investments on regional healthcare costs. Manage Sci 2018 Jun;64(6):2515-2534. [doi: 10.1287/mnsc.2017.2750]
25. Medicare and Medicaid programs; electronic health record incentive program. Final rule. Fed Regist 2010 Jul 28;75(144):44313-44588 [FREE Full text] [Medline: 20677415]
26. Blaya JA, Fraser HS, Holt B. E-health technologies show promise in developing countries. Health Aff (Millwood) 2010 Mar;29(2):244-251. [doi: 10.1377/hlthaff.2009.0894] [Medline: 20348068]
27. Garrido M, Kristensen F, Busse R, Nielsen C. Health Technology Assessment and Health Policy-making in Europe: Current Status, Challenges and Potential. Regional Office for Europe: World Health Organization; 2008:1-181.
28. Demirkan H. A smart healthcare systems framework. IT Prof 2013 Sep;15(5):38-45. [doi: 10.1109/mitp.2013.35]
29. Kim J. A review of cyber-physical system research relevant to the emerging IT trends: Industry 4.0, IoT, Big Data, and Cloud Computing. J Ind Integr Manag 2017 Nov 30;02(03):1750011. [doi: 10.1142/s2424862217500117]
30. Das S, Yaylacicegi U, Menon N. The effect of information technology investments in healthcare: a longitudinal study of its lag, duration, and economic value. IEEE Trans Eng Manage 2011 Feb;58(1):124-140. [doi: 10.1109/tem.2010.2048906]
31. Davies L, Drummond M, Papanikoloau P. Prioritising investments in health technology assessment: can we assess the potential value for money? Centre for Health Economics, University of York. 1999. URL: https://ideas.repec.org/p/chy/respap/170chedp.html [accessed 2021-07-30]
32. Thouin M, Hoffman J, Ford E. The effect of information technology investment on firm-level performance in the health care industry. Health Care Manage Rev 2008;33(1):60-68. [doi: 10.1097/01.HMR.0000304491.03147.06] [Medline: 18091445]
33. Kohli R, Devaraj S, Ow T. Does information technology investment influence a firm's market value? A case of non-publicly traded healthcare firms. MIS Q 2012;36(4):1145. [doi: 10.2307/41703502]
34. Galani C, Rutten FF. Self-reported healthcare decision-makers' attitudes towards economic evaluations of medical technologies. Curr Med Res Opin 2008 Nov;24(11):3049-3058. [doi: 10.1185/03007990802442695] [Medline: 18826747]
35. Trigeorgis L, Tsekrekos A. Real options in operations research: a review. Eur J Oper Res 2018 Oct;270(1):1-24. [doi: 10.1016/j.ejor.2017.11.055]
36. Henshall C, Schuller T. Health technology assessment, value-based decision making, and innovation. Int J Technol Assess Health Care 2013 Jul 11;29(4):353-359. [doi: 10.1017/s0266462313000378]
37. de Grood C, Raissi A, Kwon Y, Santana M. Adoption of e-health technology by physicians: a scoping review. J Multidiscip Healthc 2016 Aug;Volume 9:335-344. [doi: 10.2147/jmdh.s103881]
38. Sucharew H, Macaluso M. Progress notes: methods for research evidence synthesis: the scoping review approach. J Hosp Med 2019 Jul 01;14(7):416-418. [doi: 10.12788/jhm.3248] [Medline: 31251164]
39. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010 Sep 20;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
40. Pham M, Rajić A, Greig J, Sargeant J, Papadopoulos A, McEwen S. A scoping review of scoping reviews: advancing the approach and enhancing the consistency. Res Synth Methods 2014 Dec;5(4):371-385 [FREE Full text] [doi: 10.1002/jrsm.1123] [Medline: 26052958]

41. Lasi H, Fettke P, Kemper H, Feld T, Hoffmann M. Industry 4.0. Bus Inf Syst Eng 2014 Jun 19;6(4):239-242. [doi: 10.1007/s12599-014-0334-4]
42. Borges G, Tortorella G, Rossini M, Portioli-Staudacher A. Lean implementation in healthcare supply chain: a scoping review. J Health Organ Manag 2019 May 20;33(3):304-322. [doi: 10.1108/jhom-06-2018-0176]
43. Khan K, Ter RG, Glanville J, Sowden A, Kleijnen J. Undertaking Systematic Reviews of Research on Effectiveness: CRD's Guidance for Carrying Out or Commissioning Reviews. University of York, York: NHS Centre for Reviews and Dissemination; 2001.
44. Dreyfuss PD, Roberts TG. Making investments in medical technology: time to get real about real options. Oncologist 2011;16(12):1672-1674 [FREE Full text] [doi: 10.1634/theoncologist.2011-0345] [Medline: 22147001]
45. Grutters JP, Abrams KR, de Ruysscher D, Pijls-Johannesma M, Peters HJ, Beutner E, et al. When to wait for more evidence? Real options analysis in proton therapy. Oncologist 2011;16(12):1752-1761 [FREE Full text] [doi: 10.1634/theoncologist.2011-0029] [Medline: 22147003]
46. Marsh K, Phillips C, Fordham R, Bertranou E, Hale J. Estimating cost-effectiveness in public health: a summary of modelling and valuation methods. Health Econ Rev 2012 Sep 03;2(1):17 [FREE Full text] [doi: 10.1186/2191-1991-2-17] [Medline: 22943762]
47. Favato G, Baio G, Capone A, Marcellusi A, Saverio Mennini F. A novel method to value real options in health care: the case of a multicohort human papillomavirus vaccination strategy. Clin Ther 2013 Jul;35(7):904-914. [doi: 10.1016/j.clinthera.2013.05.003] [Medline: 23806328]
48. Drummond M, Tarricone R, Torbica A. Assessing the added value of health technologies: reconciling different perspectives. Value Health 2013;16(1 Suppl):7-13 [FREE Full text] [doi: 10.1016/j.jval.2012.10.007] [Medline: 23317646]
49. Pertile P, Forster M, La Torre D. Optimal Bayesian sequential sampling rules for the economic evaluation of health technologies. J R Stat Soc Stat Soc Series A 2014;177(2):419-438. [doi: 10.1111/rssa.12025]
50. Boydell KM, Hodgins M, Pignatiello A, Teshima J, Edwards H, Willis D. Using technology to deliver mental health services to children and youth: a scoping review. J Can Acad Child Adolesc Psychiatry 2014 May;23(2):87-99 [FREE Full text] [Medline: 24872824]
51. Kvedar J, Coye MJ, Everett W. Connected health: a review of technologies and strategies to improve patient care with telemedicine and telehealth. Health Aff (Millwood) 2014 Feb;33(2):194-199. [doi: 10.1377/hlthaff.2013.0992] [Medline: 24493760]
52. Atwood D, Larose P, Uttley R. Strategies for success in purchasing medical technology. Biomed Instrum Technol 2015;49(2):93-98. [doi: 10.2345/0899-8205-49.2.93] [Medline: 25793338]
53. Gobbi C, Hsuan J. Collaborative purchasing of complex technologies in healthcare: implications for alignment strategies. Int J Oper Prod Manag 2015;35(3):430-455. [doi: 10.1108/IJOPM-08-2013-0362]
54. Merlo G, Page K, Ratcliffe J, Halton K, Graves N. Bridging the gap: exploring the barriers to using economic evidence in healthcare decision making and strategies for improving uptake. Appl Health Econ Health Policy 2015 Jun;13(3):303-309. [doi: 10.1007/s40258-014-0132-7] [Medline: 25288052]
55. Sharma L, Chandrasekaran A, Boyer KK, McDermott CM. The impact of Health Information Technology bundles on hospital performance: an econometric study. J Oper Manag 2015 Dec 01;41(1):25-41. [doi: 10.1016/j.jom.2015.10.001]
56. 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]
57. Lavallée JF, Gray TA, Dumville J, Russell W, Cullum N. The effects of care bundles on patient outcomes: a systematic review and meta-analysis. Implement Sci 2017 Nov 29;12(1):142 [FREE Full text] [doi: 10.1186/s13012-017-0670-0] [Medline: 29187217]
58. Kim BY, Lee J. Smart devices for older adults managing chronic disease: a scoping review. JMIR Mhealth Uhealth 2017 May 23;5(5):e69 [FREE Full text] [doi: 10.2196/mhealth.7141] [Medline: 28536089]
59. Rejeb O, Pilet C, Hamana S, Xie X, Durand T, Aloui S, et al. Performance and cost evaluation of health information systems using micro-costing and discrete-event simulation. Health Care Manag Sci 2018 Jun;21(2):204-223. [doi: 10.1007/s10729-017-9402-x] [Medline: 28516345]
60. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
61. Long L, Pariyo G, Kallander K. Digital technologies for health workforce development in low- and middle-income countries: a scoping review. Glob Health Sci Pract 2018 Dec 10;6(Suppl 1):41-48 [FREE Full text] [doi: 10.9745/GHSP-D-18-00167] [Medline: 30305338]
62. Adjekum A, Blasimme A, Vayena E. Elements of trust in digital health systems: scoping review. J Med Internet Res 2018 Dec 13;20(12):e11254 [FREE Full text] [doi: 10.2196/11254] [Medline: 30545807]
63. Winters N, Langer L, Geniets A. Scoping review assessing the evidence used to support the adoption of mobile health (mHealth) technologies for the education and training of community health workers (CHWs) in low-income and

middle-income countries. BMJ Open 2018 Jul 30;8(7):e019827 [FREE Full text] [doi: 10.1136/bmjopen-2017-019827] [Medline: 30061430]

64. Baines D, Gahir IK, Hussain A, Khan AJ, Schneider P, Hasan SS, et al. A scoping review of the quality and the design of evaluations of mobile health, telehealth, smart pump and monitoring technologies performed in a pharmacy-related setting. Front Pharmacol 2018 Jul 26;9:678 [FREE Full text] [doi: 10.3389/fphar.2018.00678] [Medline: 30093859]
65. Taj F, Klein MC, van Halteren A. Digital health behavior change technology: bibliometric and scoping review of two decades of research. JMIR Mhealth Uhealth 2019 Dec 13;7(12):e13311 [FREE Full text] [doi: 10.2196/13311] [Medline: 31833836]
66. Dogba MJ, Dossa AR, Breton E, Gandonou-Migan R. Using information and communication technologies to involve patients and the public in health education in rural and remote areas: a scoping review. BMC Health Serv Res 2019 Feb 19;19(1):128 [FREE Full text] [doi: 10.1186/s12913-019-3906-7] [Medline: 30782147]
67. Loncar-Turukalo T, Zdravevski E, da Silva JM, Chouvarda I, Trajkovik V. Literature on wearable technology for connected health: scoping review of research trends, advances, and barriers. J Med Internet Res 2019 Sep 05;21(9):e14017 [FREE Full text] [doi: 10.2196/14017] [Medline: 31489843]
68. Shahid N, Rappon T, Berta W. Applications of artificial neural networks in health care organizational decision-making: a scoping review. PLoS One 2019;14(2):e0212356 [FREE Full text] [doi: 10.1371/journal.pone.0212356] [Medline: 30779785]
69. Wüller H, Behrens J, Garthaus M, Marquard S, Remmers H. A scoping review of augmented reality in nursing. BMC Nurs 2019 May 16;18(1):19 [FREE Full text] [doi: 10.1186/s12912-019-0342-2] [Medline: 31123428]
70. Chouvarda I, Maramis C, Livitckaia K, Trajkovik V, Burmaoglu S, Belani H, ENJECT Working Group 1 Network. Connected Health Services: Framework for an impact assessment. J Med Internet Res 2019 Sep 03;21(9):e14005 [FREE Full text] [doi: 10.2196/14005] [Medline: 31482857]
71. Hasselgren A, Kralevska K, Gligoroski D, Pedersen SA, Faxvaag A. Blockchain in healthcare and health sciences - a scoping review. Int J Med Inform 2020 Feb;134:104040 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.104040] [Medline: 31865055]
72. Peng C, He M, Cutrona SL, Kiefe CI, Liu F, Wang Z. Theme trends and knowledge structure on mobile health apps: bibliometric analysis. JMIR Mhealth Uhealth 2020 Jul 27;8(7):e18212 [FREE Full text] [doi: 10.2196/18212] [Medline: 32716312]
73. Ismail L, Materwala H, Karduck AP, Adem A. Requirements of health data management systems for biomedical care and research: scoping review. J Med Internet Res 2020 Jul 07;22(7):e17508 [FREE Full text] [doi: 10.2196/17508] [Medline: 32348265]
74. Wernz C, Gehrke I, Ball DR. Managerial decision-making in hospitals with real options analysis. Inf Syst E-Bus Manage 2013 Nov 23;13(4):673-691. [doi: 10.1007/s10257-013-0230-3]
75. Vassolo R, Anand J, Folta T. Non-additivity in portfolios of exploration activities: a real options-based analysis of equity alliances in biotechnology. Strat Mgmt J 2004 Sep 27;25(11):1045-1061. [doi: 10.1002/smj.414]
76. Budler M, Trkman P. The nature of management frameworks. J Manag Organ 2019 Dec 17:1-18. [doi: 10.1017/jmo.2019.83]
77. Priem RL, Butler JE. Is the resource-based "View" a useful perspective for strategic management research? Acad Manag Rev 2001 Jan;26(1):22-40. [doi: 10.5465/amr.2001.4011928]
78. Heylighen F. What makes a meme successful? Selection criteria for cultural evolution. Association Internationale de Cybernetique, Namur. 1998. URL: http://pespmc1.vub.ac.be/Papers/MemeticsNamur.html [accessed 2021-07-30]
79. Andrew N, Evans L. Approaches and frameworks for management and research in small-scale fisheries. In: Pomeroy RS, Andrew NL, editors. Small-Scale Fisheries Management: Frameworks and Approaches for the Developing World. Wallingford: CABI Publishers; 2011.
80. Schwartz M, Carroll A. Integrating and unifying competing and complementary frameworks. Bus Soc 2007 Sep 18;47(2):148-186. [doi: 10.1177/0007650306297942]
81. Cronin M, George E. The why and how of the integrative review. Organ Res Methods 2020 Jul 06:109442812093550. [doi: 10.1177/1094428120935507]
82. Snyder H. Literature review as a research methodology: an overview and guidelines. J Bus Res 2019 Nov;104:333-339. [doi: 10.1016/J.JBUSRES.2019.07.039]
83. Baden-Fuller C, Morgan MS. Business models as models. Long Range Planning 2010 Apr;43(2-3):156-171. [doi: 10.1016/j.lrp.2010.02.005]
84. DiMaggio PJ, Powell WW. The Iron Cage Revisited: Institutional isomorphism and collective rationality in organizational fields. Am Sociol Rev 1983 Apr;48(2):147. [doi: 10.2307/2095101]
85. Gastaldi L, Corso M. Smart Healthcare Digitalization: Using ICT to effectively balance exploration and exploitation within hospitals. Int J Eng Bus Manag 2012 Jan 01;4:9. [doi: 10.5772/51643]
86. Alrige M, Chatterjee S. Toward a taxonomy of wearable technologies in healthcare. In: Donnellan B, Helfert M, Kenneally J, VanderMeer D, Rothenberger M, Winter R, editors. New Horizons in Design Science: Broadening the Research Agenda. Switzerland: Springer; 2015:496-504.
87. Kulatilaka N, Trigeorgis L. The general flexibility to switch: real options revisited. Real Options and Investment Under Uncertainty: Classical Readings and Recent Contributions. 2004. URL: https://scholar.google.com/

citations?view_op=view_citation&hl=en&user=s2_bH5cAAAAJ&citation_for_view=s2_bH5cAAAAJ:Y0pCki6q_DkC [accessed 2021-07-30]
88. Kulatilaka N, Perotti EC. Strategic growth options. Manag Sci 1998 Aug;44(8):1021-1031. [doi: 10.1287/mnsc.44.8.1021]
89. Anand J, Oriani R, Vassolo RS. Alliance activity as a dynamic capability in the face of a discontinuous technological change. Organiz Sci 2010 Dec;21(6):1213-1232. [doi: 10.1287/orsc.1090.0502]
90. Linsky M, Heifetz R. Leadership on the Line, With a New Preface: Staying Alive Through the Dangers of Change. Boston: Harvard Business Review Press; 2017:1-288.
91. Salvato C, Vassolo R. The sources of dynamism in dynamic capabilities. Strat Mgmt J 2017 Nov 02;39(6):1728-1752. [doi: 10.1002/smj.2703]

Abbreviations

H4.0: Healthcare 4.0

Edited by R Kukafka; submitted 28.01.21; peer-reviewed by C Abbafati, IV George, M Behzadifar, G Marquez, D Hu; comments to author 07.04.21; revised version received 11.05.21; accepted 05.07.21; published 26.08.21.

Please cite as:
Vassolo RS, Mac Cawley AF, Tortorella GL, Fogliatto FS, Tlapa D, Narayanamurthy G
Hospital Investment Decisions in Healthcare 4.0 Technologies: Scoping Review and Framework for Exploring Challenges, Trends, and Research Directions
J Med Internet Res 2021;23(8):e27571
URL: https://www.jmir.org/2021/8/e27571
doi:10.2196/27571
PMID:34435967

©Roberto Santiago Vassolo, Alejandro Francisco Mac Cawley, Guilherme Luz Tortorella, Flavio Sanson Fogliatto, Diego Tlapa, Gopalakrishnan Narayanamurthy. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Review

Diagnostic Accuracy of Smartwatches for the Detection of Cardiac Arrhythmia: Systematic Review and Meta-analysis

Scarlet Nazarian[1*], BSc, MBBS, MRCS; Kyle Lam[1*], MA (Cantab), MB BChir, MRCS; Ara Darzi[1], PC, KBE, FRS, FMedSci, HonFREng; Hutan Ashrafian[1], PhD

[1] Imperial College London, London, United Kingdom

[*] these authors contributed equally

Corresponding Author:
Hutan Ashrafian, PhD
Imperial College London
10th Floor QEQM Building, St Mary's Hospital
Praed Street
London, W2 1NY
United Kingdom
Phone: 44 7799 871 597
Email: hutan@researchtrials.net

Abstract

Background: Significant morbidity, mortality, and financial burden are associated with cardiac rhythm abnormalities. Conventional investigative tools are often unsuccessful in detecting cardiac arrhythmias because of their episodic nature. Smartwatches have gained popularity in recent years as a health tool for the detection of cardiac rhythms.

Objective: This study aims to systematically review and meta-analyze the diagnostic accuracy of smartwatches in the detection of cardiac arrhythmias.

Methods: A systematic literature search of the Embase, MEDLINE, and Cochrane Library databases was performed in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines to identify studies reporting the use of a smartwatch for the detection of cardiac arrhythmia. Summary estimates of sensitivity, specificity, and area under the curve were attempted using a bivariate model for the diagnostic meta-analysis. Studies were examined for quality using the Quality Assessment of Diagnostic Accuracy Studies 2 tool.

Results: A total of 18 studies examining atrial fibrillation detection, bradyarrhythmias and tachyarrhythmias, and premature contractions were analyzed, measuring diagnostic accuracy in 424,371 subjects in total. The signals analyzed by smartwatches were based on photoplethysmography. The overall sensitivity, specificity, and accuracy of smartwatches for detecting cardiac arrhythmias were 100% (95% CI 0.99-1.00), 95% (95% CI 0.93-0.97), and 97% (95% CI 0.96-0.99), respectively. The pooled positive predictive value and negative predictive value for detecting cardiac arrhythmias were 85% (95% CI 0.79-0.90) and 100% (95% CI 1.0-1.0), respectively.

Conclusions: This review demonstrates the evolving field of digital disease detection. The current diagnostic accuracy of smartwatch technology for the detection of cardiac arrhythmias is high. Although the innovative drive of digital devices in health care will continue to gain momentum toward screening, the process of accurate evidence accrual and regulatory standards ready to accept their introduction is strongly needed.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42020213237; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=213237.

(J Med Internet Res 2021;23(8):e28974) doi:10.2196/28974

KEYWORDS

wearables; smartwatch; cardiac arrhythmia; atrial fibrillation; cardiology; mHealth; wearable devices; screening; diagnostics; accuracy

Introduction

Background

Cardiac arrhythmia encompasses a group of conditions in which the heart beats too quickly, too slowly, or in an irregular pattern. Significant morbidity, mortality, and financial burden are associated with cardiac rhythm abnormalities [1]. Of these cardiac rhythm abnormalities, atrial fibrillation (AF) is the most common type of cardiac arrhythmia [2], and its prevalence increases sharply with age, reaching 17.8% in a European population for those aged >85 years [3,4]. The presence of AF increases the risk of ischemic stroke by five-fold [5] and can lead to other thromboembolic events. It is well recognized that AF often remains asymptomatic, and therefore, by the time of screening, the patient may have already suffered the consequences.

Although AF is the most common type of cardiac arrhythmia, other arrhythmias, such as premature cardiac contractions, are responsible for significant symptomatic burden. Premature atrial contractions have been shown to be an independent risk factor for all strokes in a longitudinal study [6]. Similarly, a cohort study found that having premature ventricular contractions resulted in a higher rate of ischemic stroke than those without contractions [7].

Conventional screening tools, in the form of 12-lead electrocardiograms (ECGs) and ambulatory electrocardiography monitors, are often unsuccessful in detecting AF or other cardiac arrhythmias, such as bradyarrhythmias or tachyarrhythmias, because of the transient nature of episodes. The episodic and infrequent nature of cardiac arrhythmias means that they are not captured within the investigation period, making diagnosis very difficult.

Recent advances in mobile health technology and wearable electronic devices allow heart rhythm monitoring to be undertaken in real time with greater comfort, ease, and engagement [8]. Wearable devices such as smartwatches show great potential for the detection of cardiac arrhythmias. Timely diagnosis of AF ensures that management is commenced early to prevent ensuing events that impact the quality of life while also relieving the burden that this poses on the health care system.

Smartwatches have gained popularity in recent years, especially as a health tool for the detection of heart rhythms. Patients with a smartwatch can self-diagnose their heart rhythm within 30 seconds using one finger [9]. These apps use photoplethysmography (PPG) from an optical sensor to analyze the pulse rate from the wrist [10]. However, adoption of the technology by both clinicians and patients requires that these devices are accurate and provide clinically applicable information in a manner that is compatible with workflow in the health setting.

Objectives

This study aims to systematically review and meta-analyze the diagnostic accuracy of smartwatches in the detection of cardiac arrhythmias.

Methods

Overview

This review was carried out and reported in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement [11]. The review was registered at the International Prospective Register of Systematic Reviews (PROSPERO ID: CRD42020213237).

Search Strategy

A thorough literature search was performed using the Embase, MEDLINE, and Cochrane Library databases. All articles published until February 2021 were included in the study. The appropriate MeSH (Medical Subject Headings) terms and free text all field searches were performed and combined with appropriate Boolean operator terms for *arrhythmias, cardiac OR irregular pulse* OR atrial fibrillation, wearable electronic devices OR smartwatch* OR wristband*, diagnosis, computer-assisted OR diagnos**, and *detect** in Embase and Ovid in MEDLINE. Search terms in the Cochrane Library included *arrhythmias, cardiac OR atrial fibrillation OR irregular pulse* OR arrhythmia*, smartwatch* OR wearable electronic device**, and *diagnosis, computer-assisted OR detect* OR diagnos**. The full search strategy is provided in Multimedia Appendix 1.

Inclusion and Exclusion Criteria

Inclusion criteria were as follows:

- studies reporting detection of cardiac arrhythmias using smartwatches;
- studies reporting sensitivity, specificity and diagnostic accuracy; or studies with adequate information to calculate these data; and
- studies published or translated into English.

Exclusion criteria were as follows:

- studies with no original data present (eg, review article, letter);
- studies with no full text available;
- studies >20 years; and
- studies without adequate data to calculate sensitivity, specificity and diagnostic accuracy data.

Study Selection

Studies obtained from the literature search were analyzed, and duplicates were removed. Title, abstract, and full-text review were performed by 2 reviewers independently, and irrelevant studies were excluded. Disagreements were settled by consensus among the reviewers.

Data Extraction

Data were extracted onto a standard spreadsheet template. Information regarding the journal, author, study design, type of smartwatch, number of subjects, and diagnostic accuracy data (sensitivity, specificity, accuracy, positive predictive value [PPV], and negative predictive value [NPV]) was selected from each paper.

Study Quality Assessment

The Quality Assessment of Diagnostic Accuracy Studies 2 tool was used to assess the risk of bias of the included studies [12]. Each domain was classified as low risk, high risk, or unclear risk of bias.

Statistical Analysis

Summary estimates of sensitivity, specificity, and area under the curve data were attempted using a bivariate model for diagnostic meta-analysis. Independent proportions and their differences were calculated and pooled using DerSimonian and Laird random effects modeling [13]. This considered both between-study and within-study variances, which contributed to study weighting. Study-specific estimates and 95% CIs were computed and represented in forest plots. Statistical heterogeneity was determined by the I^2 statistic, where <30% was low, 30%-60% was moderate, and >60% was high. Analyses were performed using Stata version 15 (StataCorp). P values of ≤.05 were considered statistically significant.

Results

Search Results and Characteristics

The database searches identified 292 studies that matched the criteria. Duplicates were removed, and 215 studies were eligible for title and abstract screening. Following this, a full-text review was undertaken, and a total of 18 studies were included in this review. Studies that failed to satisfy the inclusion criteria were excluded, and the reasons for exclusion of these articles included wrong intervention (such as the lack of use of a smartwatch) or wrong outcomes (such as studies that did not involve the detection of cardiac arrhythmias or reports on diagnostic accuracy). The study screening and selection process is shown in Figure 1.

The studies included in this systematic review were all published between 2017 and 2021. The outcome measure in the studies was mainly AF detection but also included bradyarrhythmias, tachyarrhythmias, and premature contractions. The studies measured diagnostic accuracy using smartwatches in 424,371 subjects in total. The Apple watch was used in 7 studies, Samsung smartwatches were used in 5 studies, and the remaining studies used a Huawei, Huami, or Empatica smartwatch. One study used the Wavelet wristband. Three different types of Huawei smartwatches were used in 2 studies to assess the diagnostic accuracy [14,15].

The reference standard was an ECG in most studies in the form of a 12-lead ECG, a Holter monitor, an ECG patch, telemetry, or an internet-enabled mobile ECG. In one study, an implantable cardiac monitor was used as the standard [16]. Almost all studies, except for 2 that did not specify, used PPG-based sensors to assess pulse rate. Table 1 provides the characteristics of the included studies.

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

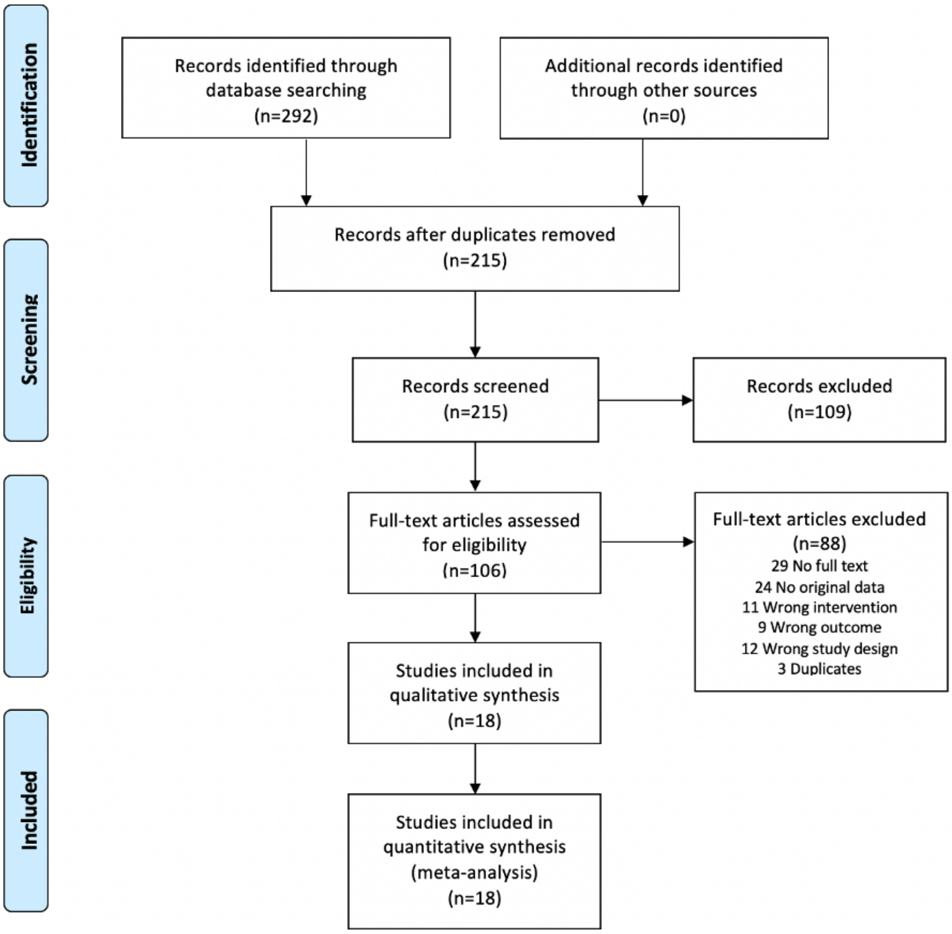

Identification
Screening
Eligibility
Included
Records identified through database searching (n=292)
Additional records identified through other sources (n=0)
Records after duplicates removed (n=215)
Records screened (n=215)
Records excluded (n=109)
Full-text articles assessed for eligibility (n=106)
Full-text articles excluded (n=88)
29 No full text
24 No original data
11 Wrong intervention
9 Wrong outcome
12 Wrong study design
3 Duplicates
Studies included in qualitative synthesis (n=18)
Studies included in quantitative synthesis (meta-analysis) (n=18)

Table 1. Characteristics of included studies on detection of cardiac arrhythmias.

Authors (year)	Primary outcome	Study design	Type of sensor	Reference standard	Research or real-life setting	Type of smartwatch	Number of subjects
Corino et al (2017) [17]	AF[a] detection	Prospective	PPG[b]	—[c]	Research	Empatica E4	70
Bumgarner et al (2018) [18]	AF detection	Prospective, nonrandomized, adjudicator blinded	—	12-lead ECG[d] (physician reviewed)	Research	Apple watch	100
Tison et al (2018) [19]	AF detection	Multinational, cohort	PPG	12-lead ECG	Research	Apple watch	1617
Wasserlauf et al (2019) [16]	AF detection	Prospective	PPG	Insertable cardiac monitor	Research	Apple watch	24
Perez et al (2019) [20]	AF detection	Prospective, single group, open label, site less, pragmatic	PPG	ECG patch	Real life	Apple watch	419,297
Zhang et al (2019) [14]	AF detection	Pilot, cohort	PPG	12-lead ECG and physical examination	Real life	Huawei Watch GT	263
						The Honor Watch (Huawei)	263
						The Honor Band4 (Huawei)	209
Ding et al (2019) [21]	AF detection	Observational	PPG	Holter monitor ECG	Research	Samsung Simband 2	40
Dorr et al (2019) [22]	AF detection	Prospective, two center, case-control	PPG	Internet-enabled mobile ECG	Research	Samsung GearFit 2	508
Bashar et al (2019) [23]	AF detection	Prospective	PPG	Holter monitor ECG	Research	Samsung Simband	20
Bashar et al (2019) [24]	AF detection	Prospective	PPG	Holter monitor ECG	Research	Samsung Simband	37
Valiaho et al (2019) [25]	AF detection	Multicenter prospective case-control	PPG	Three-lead ECG	Research	Empatica E4	213
Guo et al (2019) [15]	AF detection	Prospective	PPG	Clinical evaluation, ECG, or 24-hour Holter monitoring	Real life	Huawei Watch GT	212
						The Honor Watch (Huawei)	265
						The Honor Band4 (Huawei)	264
Chen et al (2020) [26]	AF detection	Prospective	PPG	12-lead ECG (physician reviewed)	Research	Amazfit Health Band 1S (Huami)	401
Rajakariar et al (2020) [27]	AF detection	Prospective, multicenter validation	PPG	12-lead ECG	Research	Apple watch	200
Seshadri et al (2020) [28]	AF detection	Prospective	—	Telemetry	Research	Apple watch	50
Selder et al (2020) [29]	AF detection	Observational, prospective cohort	PPG	One-lead ECG	Research	Wavelet wristband	60
Han et al (2020) [30]	Premature atrial contraction or premature ventricular contraction	Prospective	PPG	ECG patch	Research	Samsung Gear S3	2

Authors (year)	Primary outcome	Study design	Type of sensor	Reference standard	Research or real-life setting	Type of smartwatch	Number of subjects
Caillol et al (2021) [31]	AF, atrial flutter, brady arrhythmias, and tachyarrhythmias	Prospective	PPG	12-lead ECG	Research	Apple watch	256

[a] AF: atrial fibrillation.

[b] PPG: photoplethysmography.

[c] Not available.

[d] ECG: electrocardiogram.

Sensitivity, Specificity, and Accuracy

The diagnostic accuracy of the smartwatch in detecting cardiac arrhythmias was analyzed, reporting a pooled sensitivity of 100% (95% CI 0.99-1.00; Figure 2) in 17 studies with 5074 subjects and a pooled specificity of 95% (95% CI 0.93-0.97; Figure 3) in 16 studies with 5050 subjects. The sensitivity ranged from 25% (95% CI 0.14-0.36) to 100% (95% CI 1.00-1.00), whereas the specificity ranged from 68% (95% CI 0.65-0.70) to 100% (95% CI 1.00-1.00).

Of the 18 studies, 7 (39%) reported data on accuracy. Among the 1769 subjects, the pooled accuracy for arrhythmia detection was 97% (95% CI 0.96-0.99; Figure 4).

Figure 2. Pooled analysis for sensitivity of cardiac arrhythmia detection by smartwatches. Effect sizes are shown with 95% CIs. A random effects model was used. ES: effect sizes.

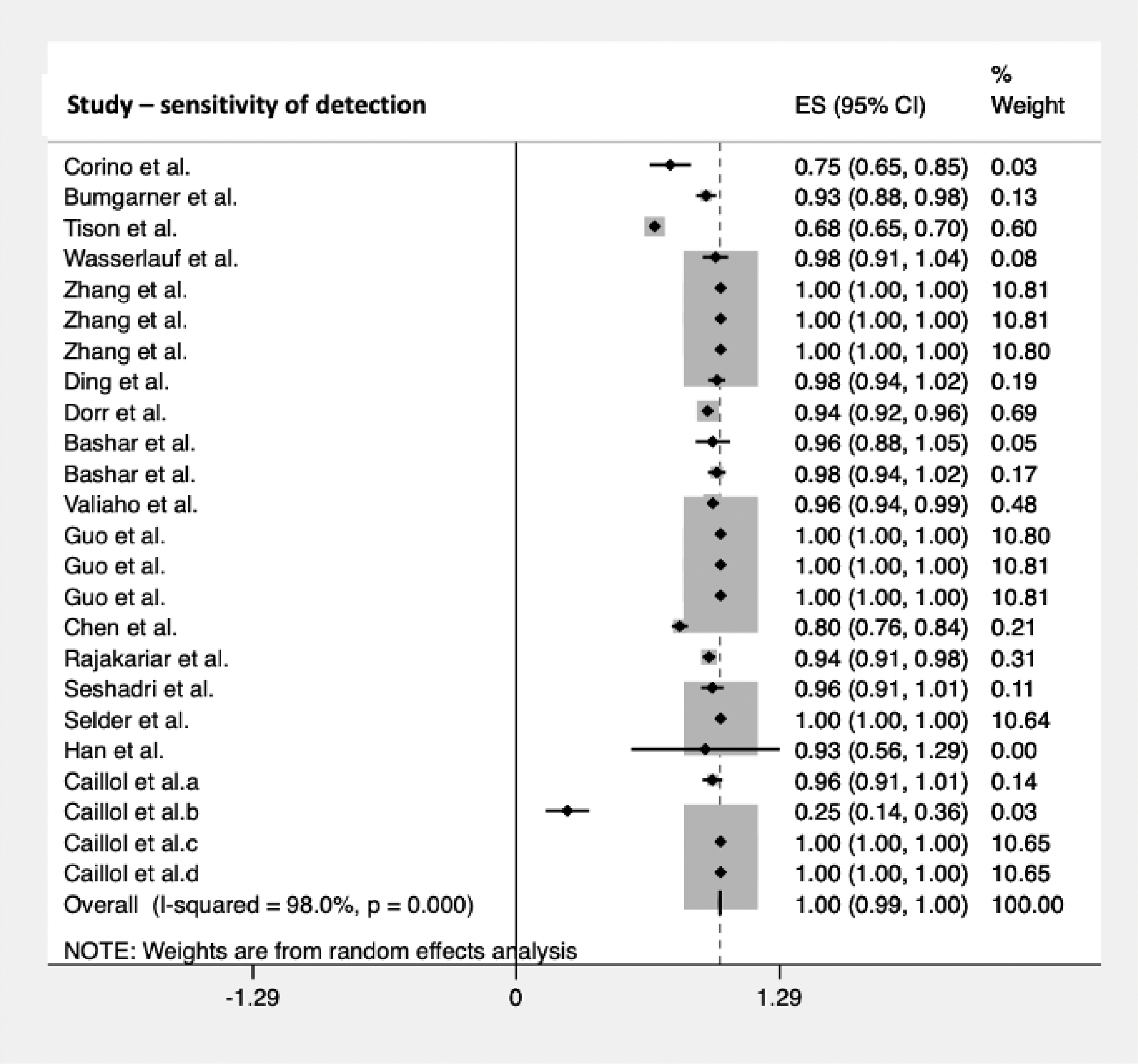


Figure 3. Pooled analysis for specificity of cardiac arrhythmia detection by smartwatches. Effect sizes are shown with 95% CIs. A random effects model was used. ES: effect sizes.

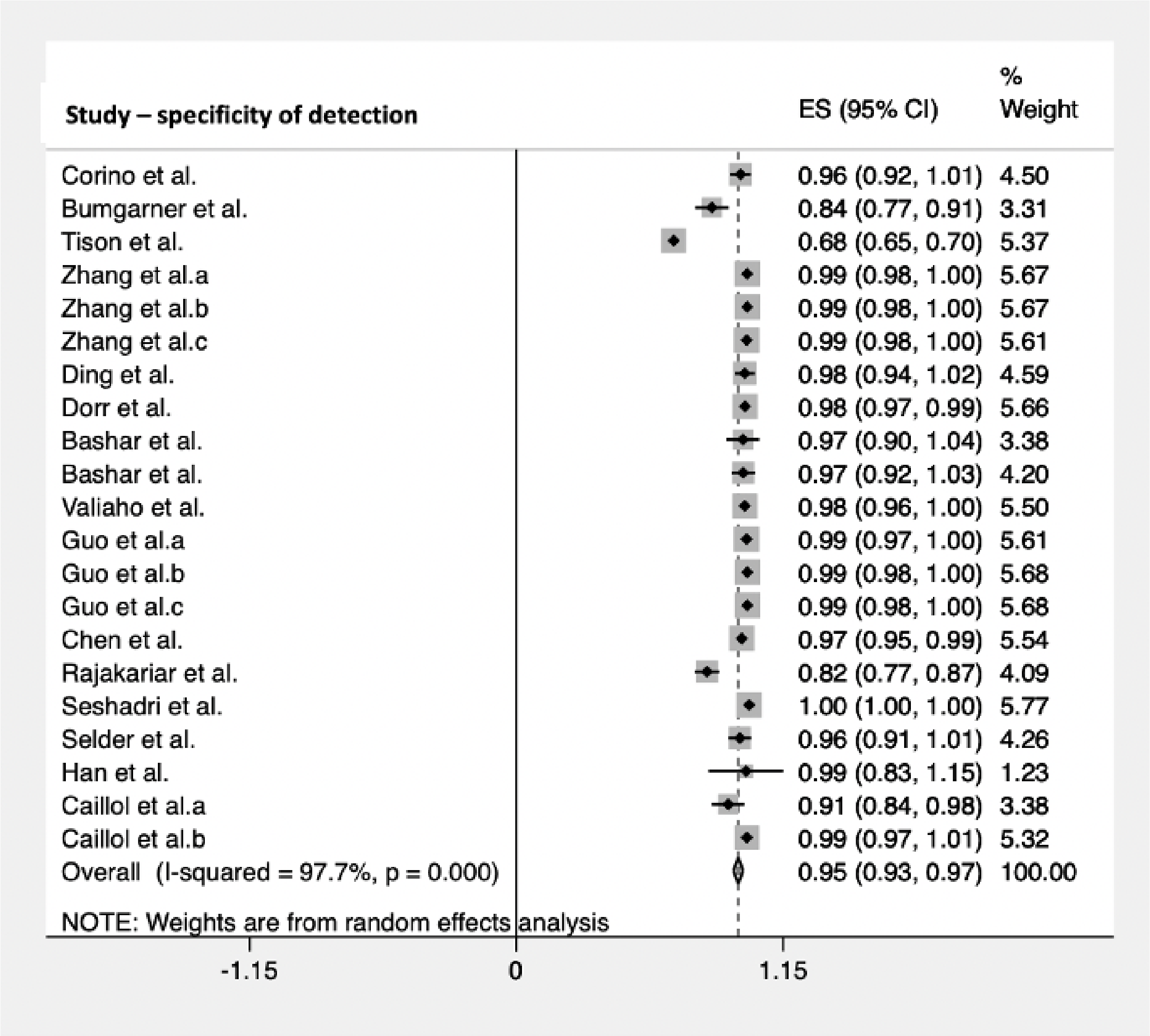
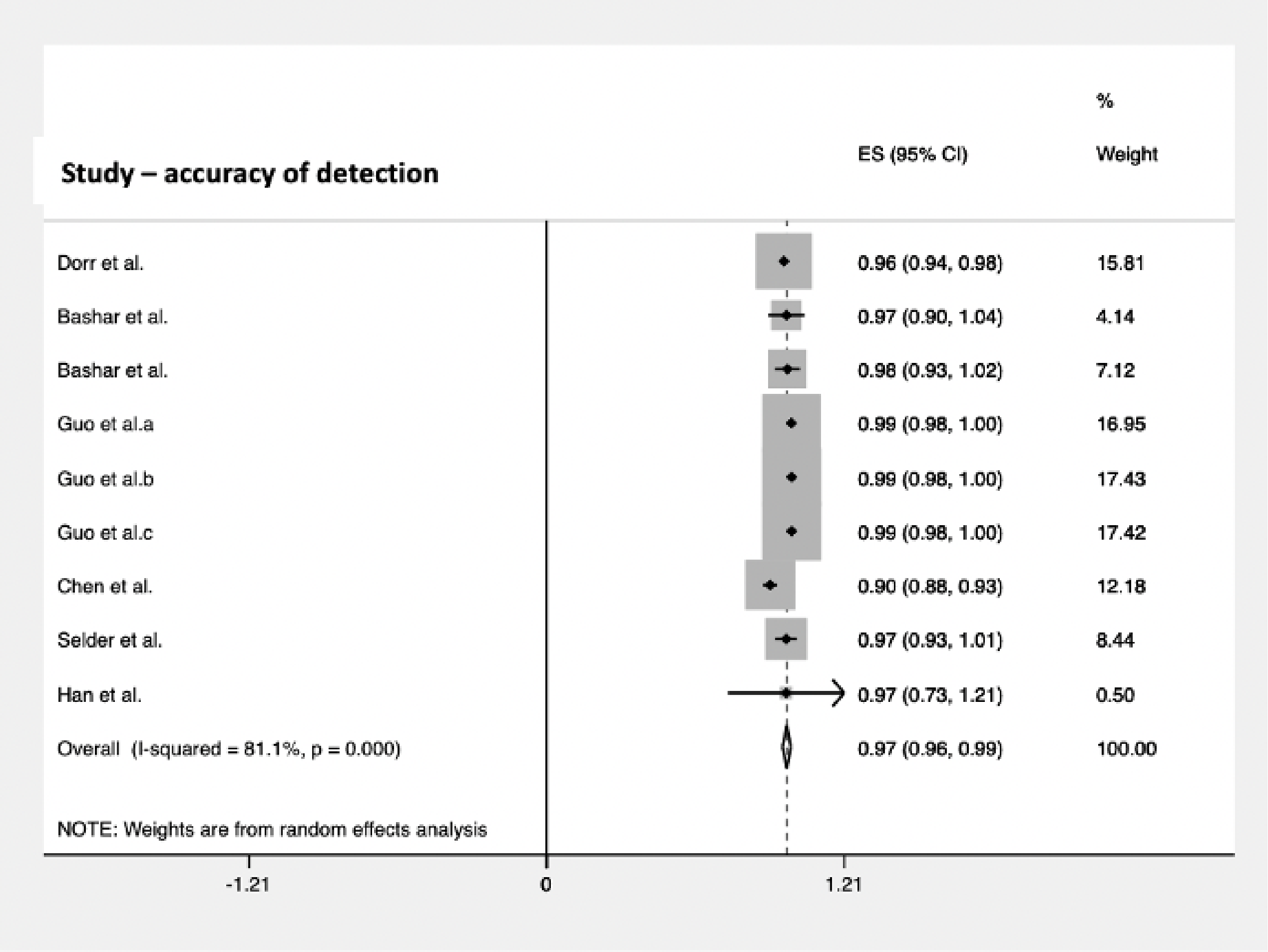


Figure 4. Pooled analysis for accuracy of cardiac arrhythmia detection by smartwatches. Effect sizes are shown with 95% CIs. A random effects model was used. ES: effect sizes.



PPV and NPV Analysis

The PPV for cardiac arrhythmia detection was assessed in 9 studies using a smartwatch. These included a total of 421,267 subjects and reported a PPV of 85% (95% CI 0.79-0.90; Figure 5). The pooled NPV was reported in 6 studies as 100% (95% CI 1.0-1.0; Figure 6), taking into consideration 3323 subjects.

Figure 5. Pooled analysis for PPV of cardiac arrhythmia detection by smartwatches. Effect sizes are shown with 95% CIs. A random effects model was used. ES: effect sizes; PPV: positive predictive value.

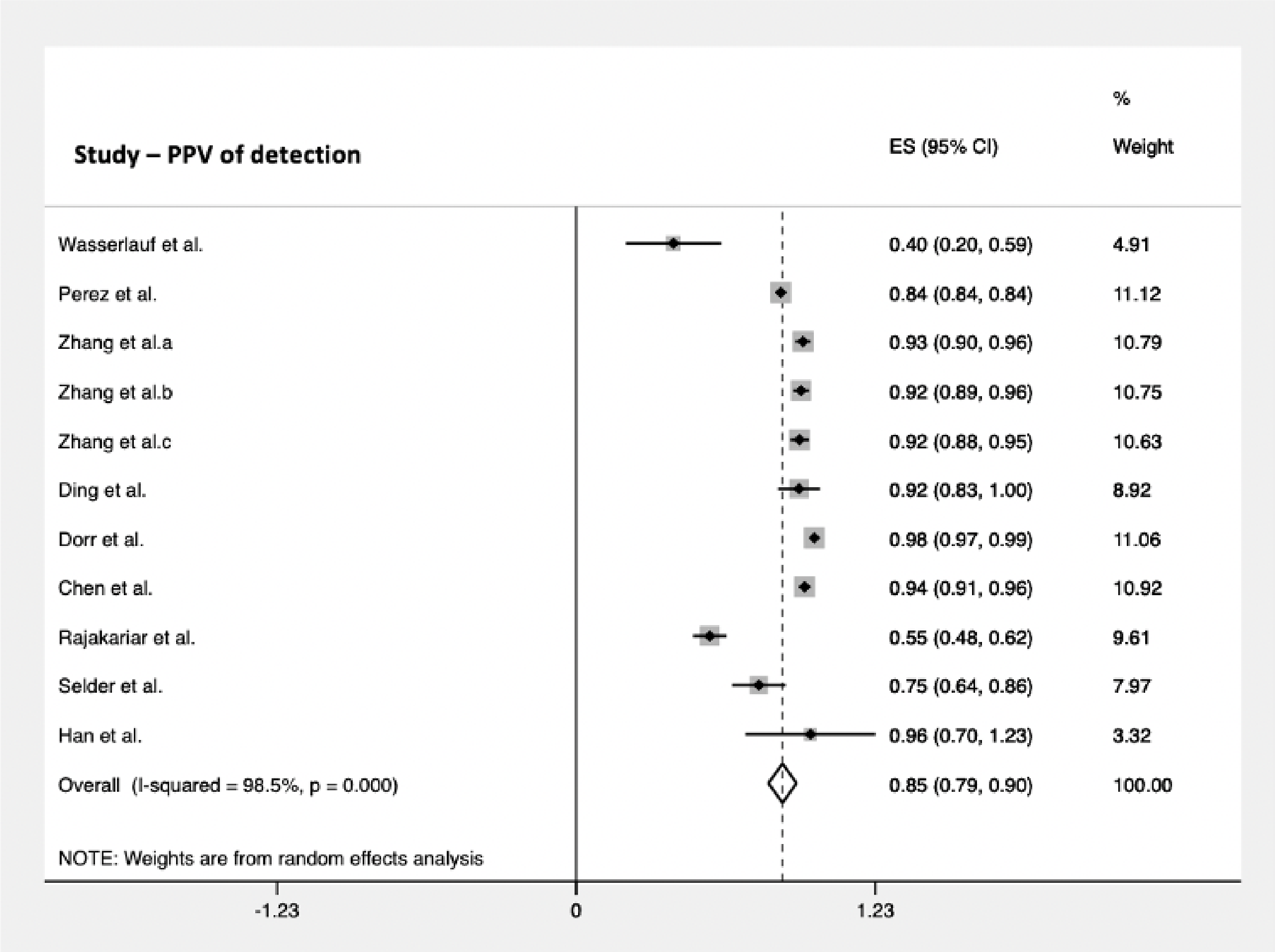
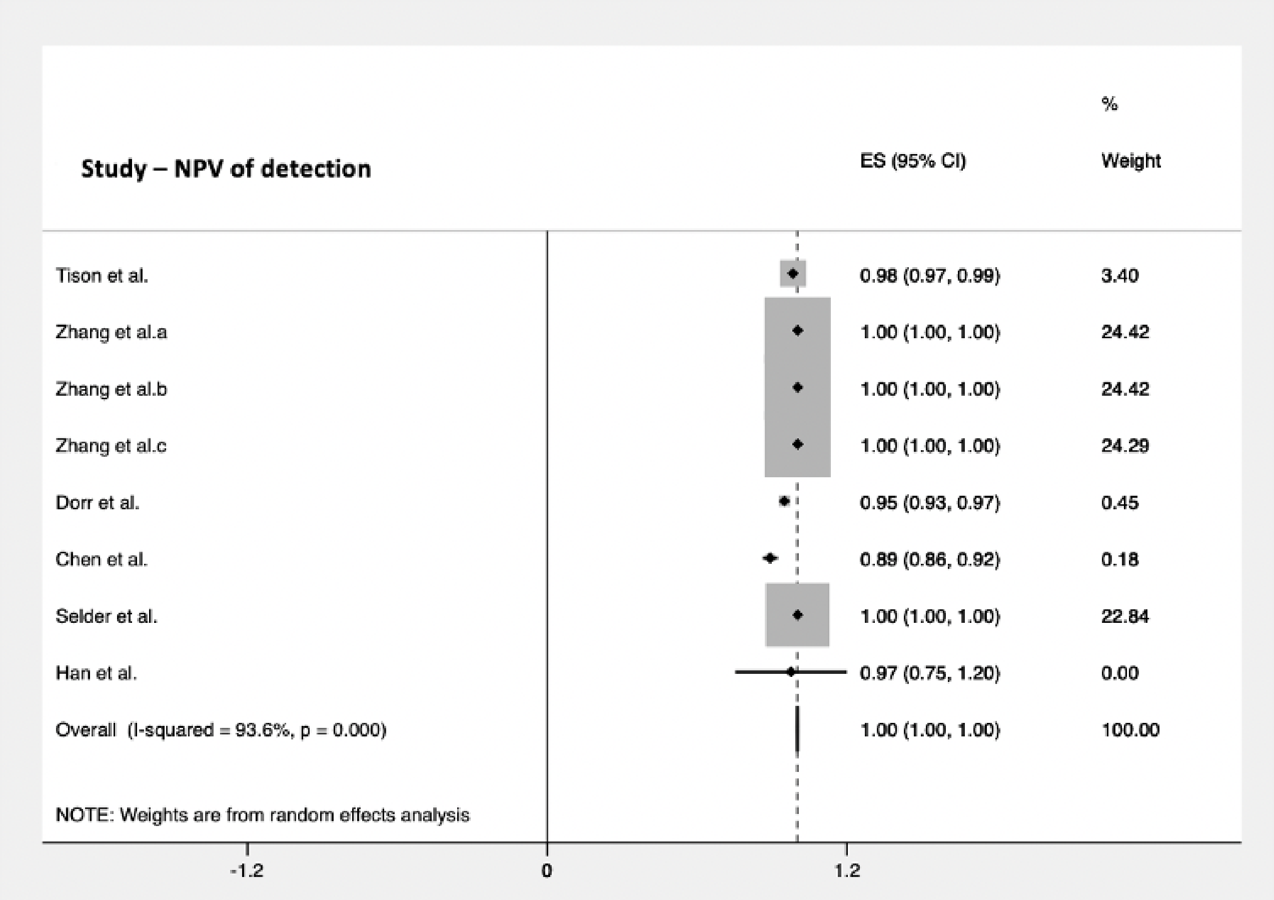


Figure 6. Pooled analysis for NPV of cardiac arrhythmia detection by smartwatches. Effect sizes are shown with 95% CIs. A random effects model was used. ES: effect sizes; NPV: negative predictive value.



Heterogeneity of Studies

There was a high degree of variation between studies assessing cardiac arrhythmia detection using a smartwatch. The heterogeneity was statistically significant when all the studies were compared ($P<.05$). The lowest variation among studies was seen when reporting the accuracy of smart devices to detect arrhythmias ($I^2=81.1\%$), whereas heterogeneity was highest in studies when assessing PPV ($I^2=98.5\%$).

Quality Assessment

The assessment of bias using the Quality Assessment of Diagnostic Accuracy Studies 2 tool for the included studies is highlighted in Multimedia Appendix 2 [14-31].

Discussion

Principal Findings

To the best of our knowledge, this systematic review and meta-analysis is the first to investigate the diagnostic accuracy of smartwatches for all cardiac arrhythmias. We have shown that the detection of cardiac arrhythmias using commercially available smartwatches is possible, with very high diagnostic accuracy. The overall sensitivity, specificity, and accuracy of these digital systems were 100%, 95%, and 97%, respectively. The pooled PPV and NPV for detecting cardiac arrhythmias were 85% and 100%, respectively. These values may offer clinicians a quantifiable appreciation for the use of smartwatches in a health care setting.

Although the aim of this study is to review the diagnostic accuracy of smartwatches in detecting cardiac arrhythmias, it is clear from the results that there are currently very few studies that assess the ability of PPG technology on smartwatches to detect non-AF arrhythmias.

Smartwatches

A wide variety of smartwatches are commercially available, and this is reflected in the diverse range of smartwatches used in these studies (Table 2). These devices range from fitness trackers to more medically oriented watches with prices between US $40 and US $1700. Although all devices use PPG sensors (Figure 7), there is diversity in functionality beyond this point. Several smartwatches are capable of recording a single-lead ECG, and others, such as the Empatica E4, have electrodermal activity sensors capable of recording sympathetic nervous system activity. The Samsung Simband is unique within these studies in that it is the only device designed for developers and is not commercially available, allowing custom adaption of sensor inclusion. Of the studies included, only the Apple Smartwatch has Food and Drug Administration (FDA) clearance for its ECG tracking functionality.

Table 2. Characteristics of smartwatches used in included studies.

Smartwatch	Company	Country	Approximate price[a], £ (US $)	Type	Photoplethysmography	Single-lead ECG[b]	Food and Drug Administration clearance ECG tracking	Electrodermal activity sensor
Apple Watch	Apple	USA	388 (531)	Watch	✓[c]	✓	✓	
Honor Watch	Honor	China	86 (117)	Watch	✓			
Huawei GT	Huawei	China	89 (122)	Watch	✓			
Gear S3	Samsung	South Korea	160 (219)	Watch	✓			
Simband	Samsung	South Korea	N/A[d]	Watch	✓	✓		✓
Honor Band	Honor	China	45 (61)	Fitness Band	✓			
Amazfit Health-band	Huami	China	33 (45)	Fitness Band	✓			
GearFit2	Samsung	South Korea	49 (67)	Fitness Band	✓			
Wavelet wrist-band	Biostrap or Wavelet Health	USA	180[e] (246)	Wristband	✓			
Empatica E4	Empatica	USA	1227[f] (1682)	Wristband	✓	✓		✓

[a] Pricing as per Amazon UK website on 22/04/2021.

[b] ECG: electrocardiogram.

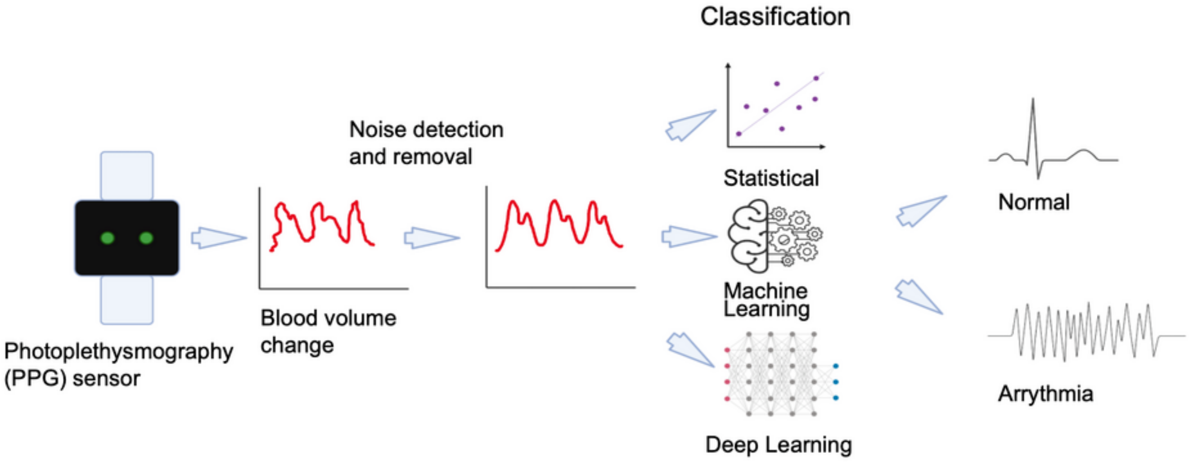
[c] Included with smartwatch.

[d] N/A: not applicable or data not available.

[e] Pricing as per Biostrap shop on 22/04/2021.

[f] Pricing as per Empatica store on 22/04/2021.

Figure 7. Overview of photoplethysmography sensor detection of arrhythmia. PPG: photoplethysmography.



The Impact of Improving AF Detection

The incidence of AF increases annually with an increase in the prevalence of risk factors, such as advancing age, obesity, hypertension, and type 2 diabetes. The challenge with detection is the ability of AF to remain asymptomatic or intermittent before eventually revealing itself. This poses a huge economic burden, accounting for 1%-2% of health care expenditure [32]. A new technology that is promising for reducing or preventing AF-related morbidity, and in doing so, addressing this burden, is welcomed. Machine learning coupled with smartwatches provides the opportunity to detect asymptomatic arrhythmias in a timely manner, allowing appropriate management to be initiated early. A recent study showed that a trained deep neural network was able to outperform single cardiologists by accurately classifying a broad range of rhythm classes and distinguishing between artifacts and arrhythmias [33]. This method could reduce the rate of misdiagnosed rhythms by digital ECG machines and improve the efficiency of expert human ECG interpretation by accurately prioritizing the most urgent conditions.

The detection of cardiac arrhythmias using smartwatches has multiple functionalities. It can be used to diagnose an abnormal rhythm, for monitoring of an arrhythmia, for example, in those with known paroxysmal AF, or for screening. Current methods of AF detection are criticized for their periodic investigative approach, during which an irregular pulse may be absent [34-36]. Using smartwatches, users can diagnose an irregular pulse by placing a finger on their device at any point. Smartwatch devices that detect cardiac arrhythmias are a simple, noninvasive, and user-friendly alternative to current ECG monitoring tools, such as 24-hour Holter monitoring or implantable cardioverter defibrillators [37,38]. The novel devices provide users with prospective information in real time, with relatively high sensitivity and specificity, as shown in our study, and are cost-effective [39]. However, the adoption of this technology by clinicians and patients requires clinically meaningful results in a manner that is compatible with the workflow of clinicians. Therefore, an optimal strategy for their implementation must be in place.

What Are the Next Steps?

Wearable devices for *wellness* are viewed as low-risk fitness monitors by the FDA, which does not apply the same stringent regulations as it would when considering medical devices. The FDA has introduced its Digital Health Precertification Program, in which companies are able to gain expedited clearance for ECG analysis and heart rate sensing software [40]. This process leads to companies producing technology that is confirmed to be *safe* but not necessarily of good quality because they have bypassed the conventional workflow for research discovery. Large clinical trials are lacking, and as a result, no expert consensus recommends screening for all occult AF [41].

Furthermore, there is insufficient evidence on the burden of smartwatch-detected AF, which would prompt further evaluation and treatment. Guidance on what the clinician is expected to do with an episode of AF detected by a smartwatch is lacking. We suggest that this should be a critical prerequisite before introducing a digital detection tool into the general population; otherwise, overdiagnosis and an expectant role of clinicians from the public to assess their device-detected condition will become an even bigger burden on the health care system. A recent study evaluating the clinical outcome of the Apple smartwatch concluded that false-positive screening results may lead to overutilization of the health care system [42]. Preparation for the problems that a new generation of smartwatch technology, which attempts to bridge the gap between disease and the health care system, brings is key.

With evolving technology in the field of health care applications, there is a move to a more personalized and *patient-centric* approach, where patients have an increasing number of tools at their disposal to assess risk and diagnose disease. Although frequent and active screening using a smartwatch is potentially feasible, few studies have examined the long-term adherence to this system. This user-involved measurement could too easily miss minimally symptomatic and brief paroxysms of arrhythmia. Long-term commitment and adherence from the user or the ability of continuous monitoring by the device is required for an accurate and worthwhile outcome.

Limitations

There are many limitations to the studies in our review. At present, most studies have assessed the use of a PPG sensor and an accompanying algorithm to detect cardiac arrhythmias. However, they have not gone further to assess the use of such systems in health care. The largest study within our systematic review did not go beyond the participants' self-reporting of an irregular pulse [20]. Several factors must be controlled to produce unbiased data that are clinically applicable. The published papers included observational and case-control studies, which did not evaluate the efficacy of smartwatch-based screening for clinical outcomes nor reflect real-life conditions [21,22]. Moreover, the sample sizes of some studies were small, with data sets of less than 50 in 5 of the papers [16,21,23,24,30]. One study had a very low sensitivity compared with others when assessing atrial flutter or tachycardia, which could likely be because of the small sample size for this group [31]. Finally, most studies were conducted in controlled research environments as opposed to a real-life setting, which may call into question the diagnostic accuracy of these smartwatches in an uncontrolled environment. Therefore, the interpretation of a sensitivity of 100%, effectively ruling out the presence of a cardiac arrhythmia with a negative result and the interpretation of an NPV of 100%, suggesting the return of no false negatives, should be interpreted with caution. The significant heterogeneity between studies is likely a result of different study settings, different patient group sizes, and different devices, based on personalized algorithms, having been used. Although the presence of this heterogeneity demands caution in interpreting our results, it also stresses the need for randomized controlled trials in this field using large data sets.

Many studies had a large proportion of data excluded because of insufficient PPG signal quality [14,18,22,24]. Some studies took place in settings where patients were supervised and provided instructions on the technique [22,27]. Thus, generalizing these findings to the *real world* could weaken the diagnostic accuracy. PPG technology recognizes the cardiac cycle by the pulsatile pattern of the change in light absorption, which reflects the volumetric alteration in the microvascular beds underneath the skin. With an accurate estimation, each episode of maximum reflected light absorption translates into an R wave. Although previous research has questioned the use of PPG sensors in darker skin, a recently published study showed no statistically significant differences in wearable heart rate measurement accuracy across skin tones [43]. However, a number of studies have shown that PPG sensors are less reliable at higher heart rates and during exercise [44,45]. As some studies in this review did not report the average heart rate of participants, it may add a level of bias to the results. In addition, PPG technology cannot detect myocardial ischemia or arrhythmias with a ventricular origin and therefore, at present, cannot completely replace 12-lead ECGs. Therefore, one must question whether the application of PPG-based sensors for cardiac arrhythmia detection is premature.

Finally, for smartwatch devices to be used as a screening tool for cardiac arrhythmias, such as AF detection, the value is highly dependent on disease prevalence. The estimated prevalence of AF in adults is between 2% and 4%. The prevalence increases with age, especially for those aged >65 years [46]. However, only 4.6% of smartwatch users in the United States are aged >65 years, and among those that are current smartwatch users, the prevalence of AF is low [47]. The studies in this review used estimated disease prevalence rates, which have been age adjusted, when assessing diagnostic accuracy values. Subjects were not limited to age groups, and as a result, some studies overestimated AF prevalence among smartwatch users [19,22]. This means that the PPV value of 85% in our review may be higher than expected for cardiac arrhythmia detection in smartwatch users. Either way, there is a high number of false positives, which leads to unnecessary anxiety among those in whom the device detects AF and may have the downstream consequences of inappropriate initiation of treatment in these patients. Treatment with anticoagulants can cause bleeding, which may be harmful. False positives may improve if the device is targeted to those most at risk of AF, but larger studies are needed to evaluate smartwatches as a tool for long-term AF screening in selected at-risk patient groups.

Regardless of the current studies, the future of health technology is undeniably advancing. Thus, measures should be taken early to ensure that such smartwatch technology supports ongoing national public health programs rather than having it run in parallel. Given the lack of recent success with the national NHS test and trace program in the United Kingdom, in which it fell short of its uptake aims when reaching contacts of people who tested positive for SARS-CoV-2 [48], the wider use of machine learning smartwatch technology should be considered in such circumstances. It may be more efficient and effective to integrate the need for health programs at the population level with existing devices. Governments should consider this, where applicable, in their decision-making processes.

Conclusions

This systematic review and meta-analysis demonstrates the evolving field of digital disease detection and the increased role of machine learning in health care. The current diagnostic accuracy of smartwatch technology for the detection of cardiac arrhythmias is high. This shift signals a new direction in the field, allowing patients to play a greater role in disease diagnosis. However, before the use of these devices as a screening tool in health care is widely adopted, more studies are needed to clearly define the ideal population for the use of these systems, as well as to help form specific guidance on the conduct of device-detected disease. Consideration should also be placed for the wider use of smartwatch technology and similar digital tools in policy making decisions by health care departments in the future. Although the innovative drive of digital devices in health care will continue to gain momentum toward screening, the process of accurate evidence accrual and regulatory standards ready to accept their introduction is strongly needed.

Acknowledgments

Infrastructure support for this research was provided by the National Institute for Health Research Imperial Biomedical Research Centre.

Conflicts of Interest

AD is Chair of the Health Security initiative at Flagship Pioneering UK Ltd. The remaining authors declare no conflicts of interest.

Multimedia Appendix 1
Search strategy.
[DOCX File , 13 KB - jmir_v23i8e28974_app1.docx]

Multimedia Appendix 2
Quality assessment.
[DOCX File , 16 KB - jmir_v23i8e28974_app2.docx]

References

1. Khurshid S, Choi SH, Weng L, Wang EY, Trinquart L, Benjamin EJ, et al. Frequency of cardiac rhythm abnormalities in a half million adults. Circ Arrhythmia Electrophysiol 2018 Jul;11(7):006273. [doi: 10.1161/circep.118.006273]
2. Benjamin EJ, Muntner P, Alonso A, Bittencourt MS, Callaway CW, Carson AP, American Heart Association Council on Epidemiology Prevention Statistics Committee Stroke Statistics Subcommittee. Heart Disease and Stroke Statistics-2019 Update: a report from the American Heart Association. Circulation 2019 Mar 05;139(10):56-528. [doi: 10.1161/CIR.0000000000000659] [Medline: 30700139]
3. Rahman F, Kwan GF, Benjamin EJ. Global epidemiology of atrial fibrillation. Nat Rev Cardiol 2014 Nov 12;11(11):639-654. [doi: 10.1038/nrcardio.2014.118] [Medline: 25113750]
4. Heeringa J, van der Kuip DA, Hofman A, Kors JA, van Herpen G, Stricker BH, et al. Prevalence, incidence and lifetime risk of atrial fibrillation: the Rotterdam study. Eur Heart J 2006 Apr;27(8):949-953. [doi: 10.1093/eurheartj/ehi825] [Medline: 16527828]
5. Wolf PA, Abbott RD, Kannel WB. Atrial fibrillation as an independent risk factor for stroke: the Framingham study. Stroke 1991 Aug;22(8):983-988. [doi: 10.1161/01.str.22.8.983] [Medline: 1866765]

6. Engström G, Hedblad B, Juul-Möller S, Tydén P, Janzon L. Cardiac arrhythmias and stroke: increased risk in men with high frequency of atrial ectopic beats. Stroke 2000 Dec;31(12):2925-2929. [doi: 10.1161/01.str.31.12.2925] [Medline: 11108750]
7. Ofoma U, He F, Shaffer ML, Naccarelli GV, Liao D. Premature cardiac contractions and risk of incident ischemic stroke. J Am Heart Assoc 2012 Sep 26;1(5):002519. [doi: 10.1161/jaha.112.002519]
8. Park H, Pei J, Shi M, Xu Q, Fan J. Designing wearable computing devices for improved comfort and user acceptance. Ergonomics 2019 Nov 03;62(11):1474-1484. [doi: 10.1080/00140139.2019.1657184] [Medline: 31437415]
9. Gil MA. Standard and precordial leads obtained with an Apple Watch. Ann Intern Med 2020 Mar 17;172(6):436-437. [doi: 10.7326/M19-2018] [Medline: 31766051]
10. Turakhia MP, Desai M, Hedlin H, Rajmane A, Talati N, Ferris T, et al. Rationale and design of a large-scale, app-based study to identify cardiac arrhythmias using a smartwatch: the Apple Heart Study. Am Heart J 2019 Jan;207:66-75 [FREE Full text] [doi: 10.1016/j.ahj.2018.09.002] [Medline: 30392584]
11. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
12. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011 Oct 18;155(8):529-536 [FREE Full text] [doi: 10.7326/0003-4819-155-8-201110180-00009] [Medline: 22007046]
13. 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]
14. Zhang H, Zhang J, Li H, Chen Y, Yang B, Guo Y, et al. Validation of single centre pre-mobile atrial fibrillation apps for continuous monitoring of atrial fibrillation in a real-world setting: pilot cohort study. J Med Internet Res 2019 Dec 03;21(12):e14909 [FREE Full text] [doi: 10.2196/14909] [Medline: 31793887]
15. Guo Y, Wang H, Zhang H, Liu T, Liang Z, Xia Y, MAFA II Investigators. Mobile photoplethysmographic technology to detect atrial fibrillation. J Am Coll Cardiol 2019 Nov 12;74(19):2365-2375. [doi: 10.1016/j.jacc.2019.08.019] [Medline: 31487545]
16. Wasserlauf J, You C, Patel R, Valys A, Albert D, Passman R. Smartwatch performance for the detection and quantification of atrial fibrillation. Circ Arrhythm Electrophysiol 2019 Jun;12(6):e006834. [doi: 10.1161/CIRCEP.118.006834] [Medline: 31113234]
17. Corino VD, Laureanti R, Ferranti L, Scarpini G, Lombardi F, Mainardi LT. Detection of atrial fibrillation episodes using a wristband device. Physiol Meas 2017 May 27;38(5):787-799. [doi: 10.1088/1361-6579/aa5dd7] [Medline: 28151434]
18. Bumgarner JM, Lambert CT, Hussein AA, Cantillon DJ, Baranowski B, Wolski K, et al. Smartwatch algorithm for automated detection of atrial fibrillation. J Am Coll Cardiol 2018 May 29;71(21):2381-2388 [FREE Full text] [doi: 10.1016/j.jacc.2018.03.003] [Medline: 29535065]
19. Tison GH, Sanchez JM, Ballinger B, Singh A, Olgin JE, Pletcher MJ, et al. Passive detection of atrial fibrillation using a commercially available smartwatch. JAMA Cardiol 2018 May 01;3(5):409-416 [FREE Full text] [doi: 10.1001/jamacardio.2018.0136] [Medline: 29562087]
20. 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]
21. Ding EY, Han D, Whitcomb C, Bashar SK, Adaramola O, Soni A, et al. Accuracy and usability of a novel algorithm for detection of irregular pulse using a smartwatch among older adults: observational study. JMIR Cardio 2019 May 15;3(1):e13850 [FREE Full text] [doi: 10.2196/13850] [Medline: 31758787]
22. Dörr M, Nohturfft V, Brasier N, Bosshard EM, Djurdjevic A, Gross S, et al. The WATCH AF Trial: SmartWATCHes for detection of atrial fibrillation. JACC Clin Electrophysiol 2019 Feb;5(2):199-208 [FREE Full text] [doi: 10.1016/j.jacep.2018.10.006] [Medline: 30784691]
23. Bashar SK, Han D, Ding E, Whitcomb C, McManus DD, Chon KH. Smartwatch based atrial fibrillation detection from photoplethysmography signals. Annu Int Conf IEEE Eng Med Biol Soc 2019 Jul;2019:4306-4309. [doi: 10.1109/EMBC.2019.8856928] [Medline: 31946820]
24. Bashar SK, Han D, Hajeb-Mohammadalipour S, Ding E, Whitcomb C, McManus DD, et al. Atrial fibrillation detection from wrist photoplethysmography signals using smartwatches. Sci Rep 2019 Oct 21;9(1):15054 [FREE Full text] [doi: 10.1038/s41598-019-49092-2] [Medline: 31636284]
25. Väliaho E, Kuoppa P, Lipponen JA, Martikainen TJ, Jäntti H, Rissanen TT, et al. Wrist band photoplethysmography in detection of individual pulses in atrial fibrillation and algorithm-based detection of atrial fibrillation. Europace 2019 Jul 01;21(7):1031-1038. [doi: 10.1093/europace/euz060] [Medline: 31505594]
26. Chen E, Jiang J, Su R, Gao M, Zhu S, Zhou J, et al. A new smart wristband equipped with an artificial intelligence algorithm to detect atrial fibrillation. Heart Rhythm 2020 May;17(5 Pt B):847-853 [FREE Full text] [doi: 10.1016/j.hrthm.2020.01.034] [Medline: 32354449]
27. Rajakariar K, Koshy A, Sajeev J, Nair S, Roberts L, Teh A. Smartwatch based arrhythmia detection: accuracy of clinician interpretation of unclassified tracings. Heart Lung Circ 2019;28:S228. [doi: 10.1016/j.hlc.2019.06.221]

28. Seshadri DR, Bittel B, Browsky D, Houghtaling P, Drummond CK, Desai MY, et al. Accuracy of Apple Watch for detection of atrial fibrillation. Circulation 2020 Feb 25;141(8):702-703. [doi: 10.1161/CIRCULATIONAHA.119.044126] [Medline: 32091929]
29. Selder J, Proesmans T, Breukel L, Dur O, Gielen W, van Rossum A, et al. Assessment of a standalone photoplethysmography (PPG) algorithm for detection of atrial fibrillation on wristband-derived data. Comput Methods Programs Biomed 2020 Dec;197:105753 [FREE Full text] [doi: 10.1016/j.cmpb.2020.105753] [Medline: 32998102]
30. Han D, Bashar SK, Mohagheghian F, Ding E, Whitcomb C, McManus DD, et al. Premature atrial and ventricular contraction detection using photoplethysmographic data from a smartwatch. Sensors (Basel) 2020 Oct 05;20(19):5683 [FREE Full text] [doi: 10.3390/s20195683] [Medline: 33028000]
31. Caillol T, Strik M, Ramirez FD, Abu-Alrub S, Marchand H, Buliard S, et al. Accuracy of a smartwatch-derived ECG for diagnosing bradyarrhythmias, tachyarrhythmias, and cardiac ischemia. Circ Arrhythmia Electrophysiol 2021 Jan;14(1):009260. [doi: 10.1161/circep.120.009260]
32. 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]
33. Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nat Med 2019 Jan 7;25(1):65-69 [FREE Full text] [doi: 10.1038/s41591-018-0268-3] [Medline: 30617320]
34. Svennberg E, Engdahl J, Al-Khalili F, Friberg L, Frykman V, Rosenqvist M. Mass screening for untreated atrial fibrillation: the STROKESTOP study. Circulation 2015 Jun 23;131(25):2176-2184. [doi: 10.1161/CIRCULATIONAHA.114.014343] [Medline: 25910800]
35. Steinhubl SR, Waalen J, Edwards AM, Ariniello LM, Mehta RR, Ebner GS, et al. Effect of a home-based wearable continuous ECG monitoring patch on detection of undiagnosed atrial fibrillation: the mSToPS randomized clinical trial. J Am Med Assoc 2018 Jul 10;320(2):146-155 [FREE Full text] [doi: 10.1001/jama.2018.8102] [Medline: 29998336]
36. Turakhia MP, Ullal AJ, Hoang DD, Than CT, Miller JD, Friday KJ, et al. Feasibility of extended ambulatory electrocardiogram monitoring to identify silent atrial fibrillation in high-risk patients: the Screening Study for Undiagnosed Atrial Fibrillation (STUDY-AF). Clin Cardiol 2015 May;38(5):285-292 [FREE Full text] [doi: 10.1002/clc.22387] [Medline: 25873476]
37. Moraes J, Rocha M, Vasconcelos G, Filho J, de Albuquerque V, Alexandria A. Advances in photopletysmography signal analysis for biomedical applications. Sensors (Basel) 2018 Jun 09;18(6):1894 [FREE Full text] [doi: 10.3390/s18061894] [Medline: 29890749]
38. Tajrishi FZ, Chitsazan M, Chitsazan M, Shojaei F, Gunnam V, Chi G. Smartwatch for the detection of atrial fibrillation. Crit Pathw Cardiol 2019 Dec;18(4):176-184. [doi: 10.1097/HPC.0000000000000192] [Medline: 31725508]
39. Lowres N, Neubeck L, Salkeld G, Krass I, McLachlan AJ, Redfern J, et al. Feasibility and cost-effectiveness of stroke prevention through community screening for atrial fibrillation using iPhone ECG in pharmacies. Thromb Haemost 2017 Nov 21;111(06):1167-1176. [doi: 10.1160/th14-03-0231]
40. Foster K, Torous J. The opportunity and obstacles for smartwatches and wearable sensors. IEEE Pulse 2019 Jan;10(1):22-25 [FREE Full text] [doi: 10.1109/mpuls.2018.2885832]
41. US Preventive Services Task Force, Curry SJ, Krist AH, Owens DK, Barry MJ, Caughey AB, et al. Screening for atrial fibrillation with electrocardiography: US Preventive Services Task Force recommendation statement. J Am Med Assoc 2018 Aug 07;320(5):478-484. [doi: 10.1001/jama.2018.10321] [Medline: 30088016]
42. Wyatt KD, Poole LR, Mullan AF, Kopecky SL, Heaton HA. Clinical evaluation and diagnostic yield following evaluation of abnormal pulse detected using Apple Watch. J Am Med Inform Assoc 2020 Jul 01;27(9):1359-1363 [FREE Full text] [doi: 10.1093/jamia/ocaa137] [Medline: 32979046]
43. Bent B, Goldstein BA, Kibbe WA, Dunn JP. Investigating sources of inaccuracy in wearable optical heart rate sensors. NPJ Digit Med 2020;3:18 [FREE Full text] [doi: 10.1038/s41746-020-0226-6] [Medline: 32047863]
44. Leader N, Dorian P, Lam J, Lee C, Woo A, Chow C. Evaluation of heart rate trackers in patients with atrial fibrillation. Can J Cardiol 2018 Oct;34(10):152-153. [doi: 10.1016/j.cjca.2018.07.174]
45. Koshy AN, Sajeev JK, Nerlekar N, Brown AJ, Rajakariar K, Zureik M, et al. Utility of photoplethysmography for heart rate estimation among inpatients. Intern Med J 2018 May 02;48(5):587-591. [doi: 10.1111/imj.13777] [Medline: 29722189]
46. Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomström-Lundqvist C, ESC Scientific Document Group. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): The Task Force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) Developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. Eur Heart J 2021 Feb 01;42(5):373-498. [doi: 10.1093/eurheartj/ehaa612] [Medline: 32860505]
47. U.S. wearable user penetration 2017, by age. Statista. 2016. URL: https://www.statista.com/statistics/739398/us-wearable-penetration-by-age/ [accessed 2020-09-24]
48. Iacobucci G. Covid-19: Is local contact tracing the answer? Br Med J 2020 Aug 17;370:m3248. [doi: 10.1136/bmj.m3248] [Medline: 32816820]

Abbreviations

AF: atrial fibrillation
ECG: electrocardiogram
FDA: Food and Drug Administration
MeSH: Medical Subject Headings
NPV: negative predictive value
PPG: photoplethysmography
PPV: positive predictive value
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by R Kukafka; submitted 20.03.21; peer-reviewed by D Filos, J Gan, L Guo; comments to author 19.04.21; revised version received 24.04.21; accepted 14.06.21; published 27.08.21.

Please cite as:
Nazarian S, Lam K, Darzi A, Ashrafian H
Diagnostic Accuracy of Smartwatches for the Detection of Cardiac Arrhythmia: Systematic Review and Meta-analysis
J Med Internet Res 2021;23(8):e28974
URL: https://www.jmir.org/2021/8/e28974
doi:10.2196/28974
PMID:34448706

©Scarlet Nazarian, Kyle Lam, Ara Darzi, Hutan Ashrafian. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Viewpoint

Every Child on the Map: A Theory of Change Framework for Improving Childhood Immunization Coverage and Equity Using Geospatial Data and Technologies

Sarah Cunard Chaney[1], MSc, MPH; Patricia Mechael[1], MHS, PhD; Nay Myo Thu[2], MSc, MBBS; Mamadou S Diallo[3], PhD; Carine Gachen[4], MSc

[1]HealthEnabled, Cape Town, South Africa
[2]UNICEF, Juba
[3]Data and Analytics Unit, Department of Analysis, Planning & Monitoring, UNICEF, New York, NY, United States
[4]Gavi, the Vaccine Alliance, Health Information Systems and Digital Health Information, Geneva, Switzerland

Corresponding Author:
Patricia Mechael, MHS, PhD
HealthEnabled
6 Wherry Road, Muizenberg
Cape Town
South Africa
Phone: 27 21 701 0766
Email: patty@healthenabled.org

Abstract

The effective use of geospatial data and technologies to collect, manage, analyze, model, and visualize geographic data has great potential to improve data-driven decision-making for immunization programs. This article presents a theory of change for the use of geospatial technologies for immunization programming—a framework to illustrate the ways in which geospatial data and technologies can contribute to improved immunization outcomes and have a positive impact on childhood immunization coverage rates in low- and middle-income countries. The theory of change is the result of a review of the state of the evidence and literature; consultation with implementers, donors, and immunization and geospatial technology experts; and a review of country-level implementation experiences. The framework illustrates how the effective use of geospatial data and technologies can help immunization programs realize improvements in the number of children immunized by producing reliable estimates of target populations, identifying chronically missed settlements and locations with the highest number of zero-dose and under-immunized children, and guiding immunization managers with solutions to optimize resource distribution and location of health services. Through these direct effects on service delivery, geospatial data and technologies can contribute to the strengthening of the overall health system with equity in immunization coverage. Recent implementation of integrated geospatial data and technologies for the immunization program in Myanmar demonstrate the process that countries may experience on the path to achieving lasting systematic improvements. The theory of change presented here may serve as a guide for country program managers, implementers, donors, and other stakeholders to better understand how geospatial tools can support immunization programs and facilitate integrated service planning and equitable delivery through the unifying role of geography and geospatial data.

(J Med Internet Res 2021;23(8):e29759) doi:10.2196/29759

KEYWORDS

geospatial data; immunization; health information systems; service delivery; equity mapping; theory; framework; children; immunization; vaccine; equity; geospatial; data; outcome; coverage; low- and middle-income; LMIC

Introduction

Maps are powerful tools for public health decision-makers to better understand the relationship between the location of populations and health system resources, indicators or predictors of health status, and their patterns over space and time. The visual power of the map is aided by modern advances in technology, computing, and handheld devices that can record the location of any place on the earth and transmit geospatial data for analysis, sharing, and use. The use of geography to analyze patterns of disease, distribution of populations, and inventories and locations of health services come together to create a catalyst for improving health systems.

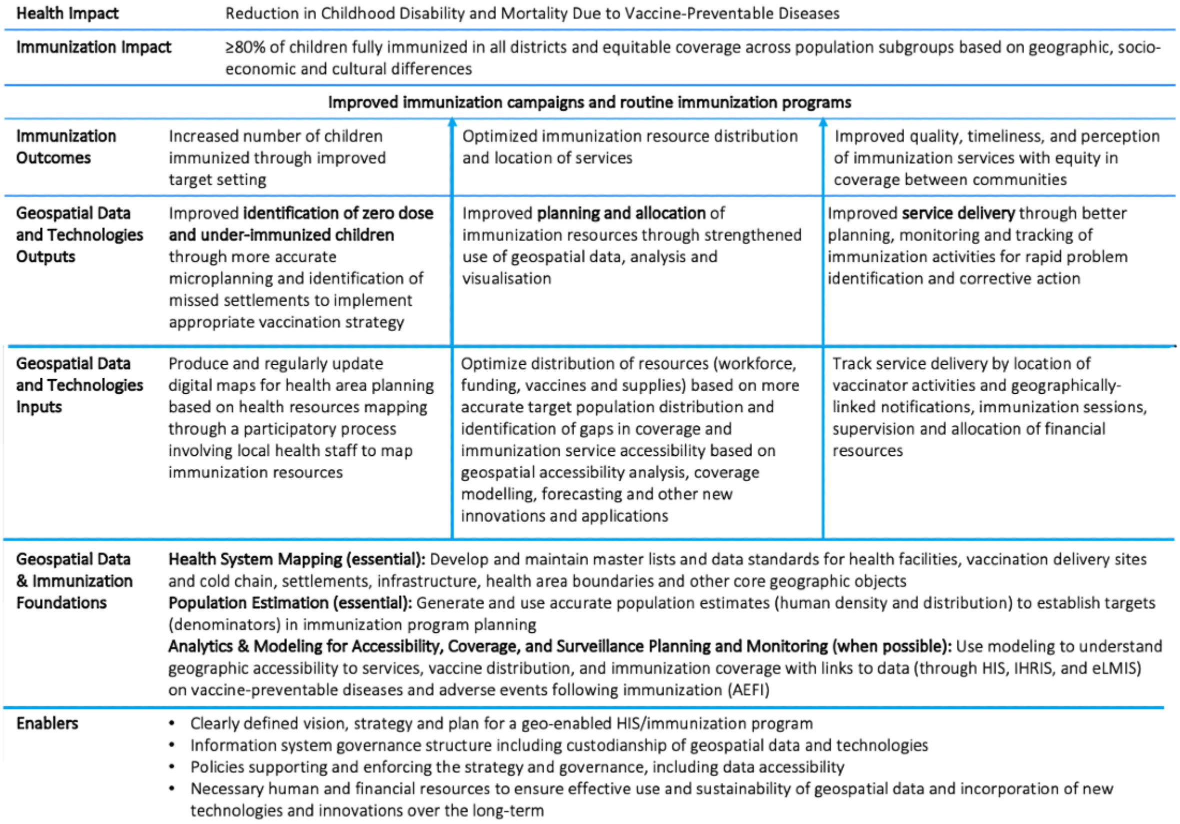
Immunization programs in low- and middle-income countries are beginning to harness digital maps and geospatial data to display and analyze complex information for program improvements [1-6]. The effective use of geospatial data can show program managers which locations have not received adequate immunization services, provide more accurate denominators, and inform what vaccination delivery strategies should be used to optimize coverage and equity. It can also improve monitoring of immunization programs.

Applications of geospatial technologies for immunization are often approached as simple solutions to system challenges without careful consideration of the greater ecosystem or planning for widespread adoption and sustainability [7]. Interventions are often deployed as pilot technology-focused projects without sustained resources or commitment to support the underlying enabling environment, human capacities, and governance systems that will contribute to a long-lasting impact on decision-making and health outcomes [8]. Gavi, the Vaccine Alliance, supports a systematic approach to understanding the range of geospatial data and technology implementation experiences to guide sustainable and effective systems and governance for improving immunization services that can reach every child with life-saving vaccines while strengthening primary health care systems [9]. Geospatial data and technology applications for immunization align with GAVI's 2021-2025 strategy and the global Immunization Agenda 2030 strategy [9,10]. In order to provide life-saving services to children who default on the vaccination schedule and "zero-dose" children who have never received a vaccine, new geo-enabled approaches to planning and delivering services are needed to expand the reach of effective vaccination for all children.

Theory of Change

Complex interventions benefit from collaborative efforts to understand the underlying series of events and changes that will lead to the desired result [11]. A *theory of change* is a process and framework to help describe this causal pathway and to support critical thinking throughout the project design, implementation, and evaluation cycle [11]. *A theory of change for the use of geospatial technologies for immunization programming* describes the potential for geospatial technologies to contribute to real-world impacts by optimizing routine immunization program design, implementation, and monitoring to reach all children with immunization services (Figure 1). It was developed as part of a collaboration between GAVI and UNICEF (United Nations Children's Emergency Fund) to review the state of evidence in the published and grey literature and through consultations with implementers, donors, immunization, and geospatial technology experts, as well as country-level implementation teams [12]. The theory of change is meant to guide future investment and planning of geospatial technologies and systems for immunization programs within a broader context of health system strengthening, to coordinate donor and partner collaboration, and optimize investments in foundations and systems for long-term sustainability and effective use of immunization data for decision-making.

Figure 1. Theory of change for the use of geospatial technologies for immunization programing (originally published and adapted from [12]), with permission from Gavi, UNICEF, and HealthEnabled.



Health Impact	Reduction in Childhood Disability and Mortality Due to Vaccine-Preventable Diseases		
Immunization Impact	≥80% of children fully immunized in all districts and equitable coverage across population subgroups based on geographic, socio-economic and cultural differences		
	Improved immunization campaigns and routine immunization programs		
Immunization Outcomes	Increased number of children immunized through improved target setting	Optimized immunization resource distribution and location of services	Improved quality, timeliness, and perception of immunization services with equity in coverage between communities
Geospatial Data and Technologies Outputs	Improved **identification of zero dose and under-immunized children** through more accurate microplanning and identification of missed settlements to implement appropriate vaccination strategy	Improved **planning and allocation** of immunization resources through strengthened use of geospatial data, analysis and visualisation	Improved **service delivery** through better planning, monitoring and tracking of immunization activities for rapid problem identification and corrective action
Geospatial Data and Technologies Inputs	Produce and regularly update digital maps for health area planning based on health resources mapping through a participatory process involving local health staff to map immunization resources	Optimize distribution of resources (workforce, funding, vaccines and supplies) based on more accurate target population distribution and identification of gaps in coverage and immunization service accessibility based on geospatial accessibility analysis, coverage modelling, forecasting and other new innovations and applications	Track service delivery by location of vaccinator activities and geographically-linked notifications, immunization sessions, supervision and allocation of financial resources
Geospatial Data & Immunization Foundations	**Health System Mapping (essential):** Develop and maintain master lists and data standards for health facilities, vaccination delivery sites and cold chain, settlements, infrastructure, health area boundaries and other core geographic objects **Population Estimation (essential):** Generate and use accurate population estimates (human density and distribution) to establish targets (denominators) in immunization program planning **Analytics & Modeling for Accessibility, Coverage, and Surveillance Planning and Monitoring (when possible):** Use modeling to understand geographic accessibility to services, vaccine distribution, and immunization coverage with links to data (through HIS, IHRIS, and eLMIS) on vaccine-preventable diseases and adverse events following immunization (AEFI)		
Enablers	• Clearly defined vision, strategy and plan for a geo-enabled HIS/immunization program • Information system governance structure including custodianship of geospatial data and technologies • Policies supporting and enforcing the strategy and governance, including data accessibility • Necessary human and financial resources to ensure effective use and sustainability of geospatial data and incorporation of new technologies and innovations over the long-term		

Evidence From Research and Implementation Experiences

Overview

Geospatial data and technologies contribute to the following three interrelated immunization outcomes in the theory of change that together strengthen immunization campaigns and routine immunization program coverage and equity:

1. Increase the number of children immunized through improved target setting
2. Optimize immunization resource distribution and location of services
3. Improve the quality, timeliness, and perception of immunization services with equity in coverage between communities

These three outcome pillars are supported by foundations and enablers in the health system and a foundation of essential data that serve to guide the collection, management, and sustainable use of geospatial data and technologies for health. The theory of change is based on evidence and implementation experiences described for each of the three expected outcomes below.

Increase the Number of Children Immunized Through Improved Target Setting

Despite years of improvements in global vaccination coverage and strengthening systems for service delivery, many children remain underimmunized or never come in contact with routine immunization programs [13]. Delivering life-saving immunization services to all children requires an enormous amount of coordination, planning, and resources; microplans are the local-level operational workplans used by immunization managers to systematically compile relevant local data, prioritize activities, maintain adequate stock, and find solutions to service delivery barriers [14]. UNICEF and the World Health Organization (WHO)'s Reach Every District (RED) strategy encourages the use of maps for local-level microplanning activities, which are traditionally hand-drawn sketches of the catchment area based on local knowledge [15]. These sketch maps are often not to scale; inaccurate or incomplete; and do not contain crucial information for microplanning such as distances, road conditions, or geographic barriers that may delay or discourage vaccinator teams from reaching remote areas during door-to-door campaign activities [3,16]. Health system data may contain overlapping borders; settlements that fall outside health boundaries; and inconsistencies in naming, spelling, and classification of service delivery units and settlements [2,4]. The planning tools and delivery strategies to reach all children with immunization services need to expand beyond the current methods to incorporate new digital tools that support local immunization managers to identify and reach areas that have been historically left off maps and microplans [17].

Children who have never received a vaccination can be clustered in settlements or neighborhoods, increasing their risk of contracting a vaccine-preventable disease without the benefit of herd immunity in their communities [18,19]. Due to a variety of socioeconomic and geographic barriers, these children are left "off the map" both literally and figuratively. Geospatial data and technologies for immunization programming can help identify these underreached communities by pinpointing the physical location of all settlements relative to the area's immunization service delivery locations. Spatially accurate maps are created through a combination of satellite images and field-based data collection to georeference and validate landmarks, inhabited settlements, and infrastructure through a participatory process with district and local immunization managers. They are then used to plan and execute realistic action plans that include outreach activities. Microplans developed with geospatial technologies and data are a cost-effective way to identify settlements missed with traditional microplanning activities that rely on hand-drawn paper maps [2,5,6,16]. With more accurate and reliable information about the locations, characteristics, and number of settlements within their catchment area, managers can plan and prioritize their activities to vaccinate more children and monitor progress both from the local and central levels.

Optimize Immunization Resource Distribution and Location of Services

Deciding how many vaccinators are needed for each catchment area, how many vaccines to send, and where to deploy fixed and outreach vaccination services depends on the number of people being served in each area, their distribution in the area, and the current unmet need for immunization services. The target population, or denominator, is often estimated from the most recent national census, adjusted each year by adding a fixed rate of growth [16]. Unfortunately, outdated census data, variation in growth rates, and population migration and mobility contribute to overestimation of the target population, leading to wasted resources, or underestimation with subsequent shortages and unvaccinated children [16,20,21]. Even with good population estimates, the location of settlements in relation to services measured by distance or travel time impact access and coverage. There is a relationship between complete and timely vaccination status and shorter distance or travel time to the nearest vaccination service, demonstrating how important the location of immunization services and geographic accessibility is for maximal immunization coverage [1,22-25]. To calculate unmet need, aggregate vaccination coverage data for the entire country or province can hide pockets of low coverage and settlements with unvaccinated children, leaving these communities vulnerable to vaccine-preventable diseases [26,27]. These data limitations impact immunization program planning and resource distribution, thereby preventing the timely delivery of life-saving vaccines to all children.

Tools and approaches that utilize geospatial technologies can help immunization managers make more targeted decisions for where and how to focus activities and resources. Precise estimates of population density and distribution for small geographic areas can be generated with a combination of satellite image data, statistical modelling, and sampled survey information to create accurate program targets for planning and monitoring purposes [28,29]. Population distribution estimates can be combined with spatial data on the location of vaccine service posts, road and transportation infrastructure, and geographic barriers to quantify the movement opportunity for people to reach existing services, inform new strategies and location of services, and prioritize outreach activities to remote communities [30,31]. Modelled vaccination coverage for small subnational units of measurement can be generated using multiple sources of data to identify pockets with low coverage and, when combined with data on other socioeconomic indicators, can help suggest solutions to overcome the social, gender-related, economic, geographic, or other factors that are preventing access to immunization services [32]. Improved granular data that is visualized to show geographic trends for local populations can help target delivery strategies and resources to increase immunization coverage in the areas that need it the most [33,34].

Improve the Quality, Timeliness, and Perception of Immunization Services With Equity in Coverage Between Communities

A number of underlying factors contribute to nonvaccination—from service delivery challenges in the immunization program to community demand, including the caregiver's perceived quality of immunization services, trust, and respect within the community [35,36]. Pockets of communities that do not receive quality and timely immunization services are susceptible to vaccine-preventable disease outbreaks. Measuring and monitoring these geographic and socioeconomic pockets of inequality is the first step toward promoting equality in coverage [37,38]. Timely and accurate data on program performance such as tracking supply and logistics, frequency of outreach services, and drop-out-rates can be used to improve the quality of services by providing entry points for supportive supervision, improve planning, identify problems, and initiate rapid corrective action for better overall service delivery [8,39]. Vaccine-preventable disease surveillance systems require rapid communication of data that facilitate feedback up and down the surveillance chain for coordinated and appropriate investigation and response [40]. In order to respond quickly to gaps and challenges, local and subnational immunization managers need to have the skills to use data that is collected accurately, transferred quickly, and presented in a way that can trigger action [8].

Mobile technologies and cellular networks provide opportunities to improve data collection, transfer, analysis, and use [41]. The combination of near real-time communication with automatic collection of accurate location data enables field-based teams of vaccinators to report on the number and location of doses delivered and any barriers encountered during immunization campaigns into an integrated dashboard where managers can monitor progress and respond appropriately to challenges and missed communities [42-45]. These daily reports of progress during campaign activities can help inform the next day's strategy or provide evidence to extend or alter activities to reach all children in the target area [43-45]. For routine immunization services, supervisors can track the progress of mobile vaccination sessions as part of a geo-enabled digital microplan to identify and respond to missed settlements and improve monitoring of the microplan implementation [3,46]. The collection of geographic information linked to reports of suspected vaccine preventable diseases can facilitate rapid and coordinated action to prevent outbreaks, identify high-risk areas that need vaccination services, and facilitate risk-mapping to predict future outbreaks [40,47]. The transparent sharing of data can promote a common understanding of expectations and challenges between vaccinators and supervisors.

Implementation Experiences: Myanmar Case Study

The current knowledge base shows that geospatial data and technology applications for immunization have the potential to stimulate programmatic improvements and increase immunization coverage. However, real-life examples of comprehensive and sustainable systems using geospatial data and technologies for immunization are rare. Myanmar provides an example of how the process of integrating geospatial data and technology for immunization microplanning validates the progression of incremental steps outlined in the theory of change.

In 2016, the national immunization program in Myanmar undertook a review as part of a health system commitment to creating a geo-enabled health information system. The assessment uncovered gaps in immunization coverage for children living in geographically and socially hard-to-reach communities, such as migrant worker settlements, remote villages, ethnic minority communities, and conflict-affected areas [48]. The local-level operational immunization workplans lacked reliable population information, and boundaries were out of date. This limited the ability of health workers to plan and undertake the daily logistics of immunization service delivery. In response to these gaps in coverage, the program took steps to support the microplanning process with geospatial data and technologies.

A phased pilot approach began in late 2017 in one township to begin building foundations, local capacity, and standard procedures and to demonstrate the benefits of using geospatial data and technologies for local-level immunization microplanning [48]. Subsequent expansion to a larger region in 2018 built on the foundations and lessons learned from the first pilot, as well as made improvements in the processes and implementation approach. Each expansion phase to a new area lasted 6 to 9 months to ensure that local capacity and systems were strengthened along the way.

The field implementation process created an up-to-date geo-referenced master list of facilities, settlements, and health area boundaries. A master list establishes a standardized, complete, up-to-date, and uniquely coded list of all features essential to the delivery of immunization services. Through this collaborative process, standard definitions were established for the geographic objects relevant to the microplanning process (eg, vaccination sites, facilities, and communities), and procedures were established for standard data collection. Every location where people lived, including temporary migrant settlements, were identified, defined, and included in the master list. Satellite images aided in settlement identification and catchment area delineation. Health workers were important stakeholders in the process to validate and review the maps and make necessary adjustments to their immunization microplans based on available transportation routes, distances, and geographic features in coordination with their supervisors. Online and printed maps showing accurate spatial relationships between key immunization assets and communities were produced and made available for national immunization program staff to plan vaccination campaigns and routine service delivery.

The interim results from Myanmar's phased implementation approach include immediate effects of the collaborative process, map production, and distribution. With settlements and communities well defined, including characteristics and locations of temporary settlements, health workers were able to include these previously overlooked populations in their immunization microplans. The addition of missed settlements improved target population estimates, allowing for improvements in service delivery planning. The transparency and sharing of microplans and maps enabled supervisors to provide better support to health workers and encouraged accountability at all levels. Health officials were able to see the need for expanded health facility distribution with a clear visualization and accurate distances displayed in new microplanning maps.

These experiences validate the expected outputs for the integration of geospatial data and technologies in the theory of change (Table 1). Myanmar's process of integrating geospatial data and technologies for immunization microplanning demonstrates how the complex challenge of delivering effective vaccinations to every child in countries with underlying health system challenges can benefit from these applications. Based on the implementation experiences in Myanmar's program, it seems likely that continued expansion and improvements in the geo-enablement of their immunization program will lead to the desired immunization outcomes and overall expanded coverage as the theory of change suggests.

Table 1. Summary of geo-enabled microplanning implementation results from the Myanmar Central Expanded Program on Immunization.

Myanmar's geo-enabled microplanning experiences	Corresponding geospatial data and technology theory of change output
• Settlements that were previously missed are defined, identified, and included in the microplan • Visualization of accurate geospatial relationships in catchment areas serve as a tool to plan vaccination sessions	Improved identification of zero-dose and underimmunized children through more accurate microplanning and identification of missed settlements to implement appropriate vaccination strategy
• Target population denominator is closer to actual community density and distribution • Standardized definitions and categorization of settlements and immunization resources help streamline planning process • Visualization serves as an advocacy tool to demonstrate to senior health officials the need for improvements in the equitable distribution and allocation of resources	Improved planning and allocation of immunization resources through strengthened use of geospatial data, analysis, and visualization
• Enhanced geo-enabled microplanning process encourages accountability of health workers and supervisors with transparency and shared expectations and service delivery plans	Improved service delivery through better planning, monitoring, and tracking of immunization activities for rapid problem identification and corrective action

Additional lessons from Myanmar's experience reinforce the importance of the enabling environment and foundations in the theory of change, built on UNICEF's guidelines and detailed approach to support the enabling environment for geospatial data and technologies in immunization programs [49]. The high-level commitment to transition to a geo-enabled national health information system in the Department of Public Health in Myanmar, with support and technical guidance from the WHO, GAVI, UNICEF, and the Health GeoLab Collaborative (a center of excellence for the Asia-Pacific region) laid a strong foundation of advocacy, governance, policies, and capacity for the management and use of geospatial data and technologies in the broader health sector [50]. The geo-enhanced microplanning process was further supported by the creation of common master lists for the geographic objects essential to the immunization program, an element recommended in the theory of change as an essential foundation to the sustained and effective use of geospatial data and technologies for immunization programs. A commitment to improving the supportive environment with a dedicated plan and resources to address needs and gaps in these enablers and foundations will promote the sustainable and effective use of geospatial data and technologies and the application of future geospatial innovations for immunization programs.

Discussion: Applying the Theory of Change

As more immunization programs begin to incorporate geospatial data and technologies to help achieve and measure improvements in equitable immunization service delivery, the *theory of change for the use of geospatial technologies for immunization programming* can guide discussions, decision-making, and consensus building for investment, development, and coordination. The theory of change represents a thought process aimed at understanding the underlying sequence of events that can contribute to sustained and effective improvements and should be considered a roadmap that is subject to change, improvements, and fine-tuning as more country-level experiences bring insights into best practices and real-world challenges. The three pillars and supportive foundations and enablers can help initiate conversations and identify needs and gaps in country immunization programs to make sound decisions for short-term and long-term planning and contribute to improving the broader health system through shared geospatial data, technologies, and resources.

The theory of change may also serve as a framework for operational research and evaluations by suggesting quantifiable research objectives that will contribute to the evidence base and help clarify the relationships and determinants of effective application and use of geospatial data and technologies. The use of geospatial data and technologies within immunization programs can improve not only the systematic collection and use of quality and transparent data for programming but also for measuring improvements and incremental achievements throughout the project cycle.

The framework presented here is grounded in lessons from a handful of implementation experiences and existing evidence from the literature. As more countries gain practical experiences in integrating geospatial data and technologies into national immunization programs, best practices will suggest improvements to this theory of change and will help guide other programs on the incremental steps, foundations, planning, and budgeting recommendations that contribute to the sustainable integration of spatial data for immunization programming. A number of global and regional centers are developing and testing practical guidance and also providing technical support, resources, and training to help national programs apply geospatial data and technologies for immunization and other health systems [49,51,52].

Conclusions

Effective data use will be necessary to make additional gains in global immunization coverage. Technology can help improve the collection, visualization, and use of data to detect and address inequalities in coverage [8,38]. However, the quality and value of immunization data ultimately depends on the people who are collecting, analyzing, and using the data, not just the technology they are using [8]. Geospatial data and technologies are a means to an end. They can strengthen data-driven decision-making if they are aligned with immunization outcomes in ways that address program needs and reinforce people’s confidence and trust in the resulting data products and analyses. Optimizing the deployment of immunization services to make them accessible for newly identified communities will pave the way for anchoring primary health care services in underserved areas. A focus on investing in and building sustainable and equitable health and immunization systems with strong leadership and capacity to use the geospatial tools and technologies that are appropriate for each country program will be critical for delivering life-saving vaccines to all children.

Acknowledgments

The authors thank the Monitoring an Evaluation team at the GAVI Secretariat and all the key informants for providing guidance on the original landscape and theory of change. Special thanks to Rocco Panceira at UNICEF for providing insights and suggestions on the manuscript. The research was supported by Gavi, the Vaccine Alliance (contract no. M&E88730520).

Authors' Contributions

SCC led landscape analysis, evidence review, cases study development, and primary author of paper. PM led framing and co-authoring of the landscape; led development of theory of change; and contributed to the structure, review, and revision of the manuscript. NMT provided content for the case study on Myanmar experiences and performed an overall review of the manuscript. MSD provided technical guidance and reviewed and revised the manuscript. CG provided guidance, framing, and input for the landscape and theory of change, as well as technical guidance, review, and revision of the manuscript.

Conflicts of Interest

None declared.

References

1. Joseph NK, Macharia PM, Ouma PO, Mumo J, Jalang'o R, Wagacha PW, et al. Spatial access inequities and childhood immunisation uptake in Kenya. BMC Public Health 2020 Sep 15;20(1):1407 [FREE Full text] [doi: 10.1186/s12889-020-09486-8] [Medline: 32933501]
2. Dougherty L, Abdulkarim M, Mikailu F, Tijani U, Owolabi K, Gilroy K, et al. From paper maps to digital maps: enhancing routine immunisation microplanning in Northern Nigeria. BMJ Glob Health 2019;4(Suppl 5):e001606 [FREE Full text] [doi: 10.1136/bmjgh-2019-001606] [Medline: 31321093]
3. Gammino V, Nuhu A, Chenoweth P, Manneh F, Young RR, Sugerman DE, et al. Using geographic information systems to track polio vaccination team performance: pilot project report. J Infect Dis 2014 Nov 01;210 Suppl 1:S98-101. [doi: 10.1093/infdis/jit285] [Medline: 25316882]
4. Rosencrans L, Sume G, Kouontchou J, Voorman A, Anokwa Y, Fezeu M, et al. Mapping for health in Cameroon: polio legacy and beyond. J Infect Dis 2017 Jul 01;216(suppl_1):S337-S342 [FREE Full text] [doi: 10.1093/infdis/jix008] [Medline: 28838181]
5. Barau I, Zubairu M, Mwanza M, Seaman VY. Improving polio vaccination coverage in Nigeria through the use of geographic information system technology. J Infect Dis 2014 Nov 01;210 Suppl 1:S102-S110. [doi: 10.1093/infdis/jiu010] [Medline: 25316823]
6. Kamadjeu R. Tracking the polio virus down the Congo River: a case study on the use of Google Earth™ in public health planning and mapping. Int J Health Geogr 2009;8(1):4. [doi: 10.1186/1476-072x-8-4]

7. Huang F, Blaschke S, Lucas H. Beyond pilotitis: taking digital health interventions to the national level in China and Uganda. Global Health 2017 Jul 31;13(1):49 [FREE Full text] [doi: 10.1186/s12992-017-0275-z] [Medline: 28756767]
8. Scobie HM, Edelstein M, Nicol E, Morice A, Rahimi N, MacDonald NE, SAGE Working Group on ImmunizationSurveillance Data QualityUse. Improving the quality and use of immunization and surveillance data: Summary report of the Working Group of the Strategic Advisory Group of Experts on Immunization. Vaccine 2020 Oct 27;38(46):7183-7197 [FREE Full text] [doi: 10.1016/j.vaccine.2020.09.017] [Medline: 32950304]
9. Gavi, the Vaccine Alliance strategy 2021-2025. Gavi, the Vaccine Alliance. 2021. URL: https://www.gavi.org/sites/default/files/board/minutes/2019/Gavi%20strategy%202021-2025%20one-pager.pdf [accessed 2021-02-28]
10. Cherian T, Hwang A, Mantel C, Veira C, Malvolti S, MacDonald N, et al. Global Vaccine Action Plan lessons learned III: monitoring and evaluation/accountability framework. Vaccine 2020 Jul 14;38(33):5379-5383 [FREE Full text] [doi: 10.1016/j.vaccine.2020.05.028] [Medline: 32430149]
11. Vogel I. Review of the use of Theory of Change in international development. UK Department of International Development. 2012. URL: https://assets.publishing.service.gov.uk/media/57a08a5ded915d3cfd00071a/DFID_ToC_Review_VogelV7.pdf [accessed 2021-01-10]
12. Gavi, The Vaccine Alliance, in Collaboration with UNICEF. 2020 Sep. URL: https://www.gavi.org/sites/default/files/document/2020/GIS-and-Immunisation-Landscape_EN.pdf [accessed 2021-01-11]
13. MacDonald N, Mohsni E, Al-Mazrou Y, Kim Andrus J, Arora N, Elden S, et al. Global vaccine action plan lessons learned I: recommendations for the next decade. Vaccine 2020 Jul 14;38(33):5364-5371 [FREE Full text] [doi: 10.1016/j.vaccine.2020.05.003] [Medline: 32563607]
14. Reaching Every District (RED) - a guide to increasing coverage and equity in all communities in the African Region. World Health Organization. 2017. URL: https://www.afro.who.int/publications/reaching-every-district-red-guide-increasing-coverage-and-equity-all-communities [accessed 2021-01-11]
15. Microplanning for immunization service delivery using the Reaching Every District (RED) strategy. World Health Organization. 2009. URL: https://apps.who.int/iris/handle/10665/70450 [accessed 2021-01-11]
16. Ali D, Levin A, Abdulkarim M, Tijjani U, Ahmed B, Namalam F, et al. A cost-effectiveness analysis of traditional and geographic information system-supported microplanning approaches for routine immunization program management in northern Nigeria. Vaccine 2020 Feb 05;38(6):1408-1415 [FREE Full text] [doi: 10.1016/j.vaccine.2019.12.002] [Medline: 31924428]
17. Berkley S. Immunization needs a technology boost. Nature 2017 Nov 16;551(7680):273. [doi: 10.1038/d41586-017-05923-8] [Medline: 29144479]
18. Wiysonge C, Uthman O, Ndumbe P, Hussey GD. Individual and contextual factors associated with low childhood immunisation coverage in sub-Saharan Africa: a multilevel analysis. PLoS One 2012;7(5):e37905 [FREE Full text] [doi: 10.1371/journal.pone.0037905] [Medline: 22662247]
19. Herliana P, Douiri A. Determinants of immunisation coverage of children aged 12-59 months in Indonesia: a cross-sectional study. BMJ Open 2017 Dec 22;7(12):e015790 [FREE Full text] [doi: 10.1136/bmjopen-2016-015790] [Medline: 29275336]
20. Ghiselli M, Wilson I, Kaplan B, Waziri N, Sule A, Ayanleke H, et al. Comparison of micro-census results for Magarya Ward, Wurno local government area of Sokoto state, Nigeria, with other sources of denominator data. Data (Basel) 2019;4(1):20 [FREE Full text] [doi: 10.3390/data4010020] [Medline: 30956970]
21. Alegana VA, Atkinson PM, Pezzulo C, Sorichetta A, Weiss D, Bird T, et al. Fine resolution mapping of population age-structures for health and development applications. J R Soc Interface 2015 Apr 06;12(105):20150073 [FREE Full text] [doi: 10.1098/rsif.2015.0073] [Medline: 25788540]
22. Sasaki S, Igarashi K, Fujino Y, Comber AJ, Brunsdon C, Muleya CM, et al. The impact of community-based outreach immunisation services on immunisation coverage with GIS network accessibility analysis in peri-urban areas, Zambia. J Epidemiol Community Health 2011 Dec;65(12):1171-1178. [doi: 10.1136/jech.2009.104190] [Medline: 20961876]
23. Metcalf CJE, Tatem A, Bjornstad ON, Lessler J, O'reilly K, Takahashi S, et al. Transport networks and inequities in vaccination: remoteness shapes measles vaccine coverage and prospects for elimination across Africa. Epidemiol. Infect 2014 Aug 14;143(7):1457-1466. [doi: 10.1017/s0950268814001988]
24. Okwaraji YB, Mulholland K, Schellenberg J, Andarge G, Admassu M, Edmond KM. The association between travel time to health facilities and childhood vaccine coverage in rural Ethiopia. A community based cross sectional study. BMC Public Health 2012 Jun 22;12:476 [FREE Full text] [doi: 10.1186/1471-2458-12-476] [Medline: 22726457]
25. Hierink F, Okiro E, Flahault A, Ray N. The winding road to health: A systematic scoping review on the effect of geographical accessibility to health care on infectious diseases in low- and middle-income countries. PLoS One 2021;16(1):e0244921 [FREE Full text] [doi: 10.1371/journal.pone.0244921] [Medline: 33395431]
26. Takahashi S, Metcalf CJE, Ferrari MJ, Tatem AJ, Lessler J. The geography of measles vaccination in the African Great Lakes region. Nat Commun 2017 May 25;8:15585 [FREE Full text] [doi: 10.1038/ncomms15585] [Medline: 28541287]
27. Utazi CE, Thorley J, Alegana VA, Ferrari MJ, Takahashi S, Metcalf CJE, et al. High resolution age-structured mapping of childhood vaccination coverage in low and middle income countries. Vaccine 2018 Mar 14;36(12):1583-1591 [FREE Full text] [doi: 10.1016/j.vaccine.2018.02.020] [Medline: 29454519]

28. Bharti N, Djibo A, Tatem AJ, Grenfell BT, Ferrari MJ. Measuring populations to improve vaccination coverage. Sci Rep 2016 Oct 05;5:34541 [FREE Full text] [doi: 10.1038/srep34541] [Medline: 27703191]
29. Tatem A. Mapping the denominator: spatial demography in the measurement of progress. Int Health 2014 Sep;6(3):153-155 [FREE Full text] [doi: 10.1093/inthealth/ihu057] [Medline: 25125576]
30. Blanford JI, Kumar S, Luo W, MacEachren AM. It's a long, long walk: accessibility to hospitals, maternity and integrated health centers in Niger. Int J Health Geogr 2012;11(1):24-15. [doi: 10.1186/1476-072x-11-24]
31. Ray N, Ebener S. AccessMod 3.0: computing geographic coverage and accessibility to health care services using anisotropic movement of patients. Int J Health Geogr 2008;7(1):63. [doi: 10.1186/1476-072x-7-63]
32. Cutts F, Dansereau E, Ferrari M, Hanson M, McCarthy KA, Metcalf CJE, et al. Using models to shape measles control and elimination strategies in low- and middle-income countries: A review of recent applications. Vaccine 2020 Jan 29;38(5):979-992 [FREE Full text] [doi: 10.1016/j.vaccine.2019.11.020] [Medline: 31787412]
33. Mosser J, Gagne-Maynard W, Rao P, Osgood-Zimmerman A, Fullman N, Graetz N, et al. Mapping diphtheria-pertussis-tetanus vaccine coverage in Africa, 2000–2016: a spatial and temporal modelling study. The Lancet 2019 May;393(10183):1843-1855. [doi: 10.1016/s0140-6736(19)30226-0]
34. Shikuku D, Muganda M, Amunga S, Obwanda EO, Muga A, Matete T, et al. Door-to-door immunization strategy for improving access and utilization of immunization services in hard-to-reach areas: a case of Migori County, Kenya. BMC Public Health 2019 Aug 07;19(1):1064 [FREE Full text] [doi: 10.1186/s12889-019-7415-8] [Medline: 31391028]
35. Streefland PH, Chowdhury AM, Ramos-Jimenez P. Quality of vaccination services and social demand for vaccinations in Africa and Asia. Bull World Health Organ 1999;77(9):722-730 [FREE Full text] [Medline: 10534895]
36. Favin M, Steinglass R, Fields R, Banerjee K, Sawhney M. Why children are not vaccinated: a review of the grey literature. Int Health 2012 Dec;4(4):229-238. [doi: 10.1016/j.inhe.2012.07.004] [Medline: 24029668]
37. Inequality monitoring in immunization: a step-by-step manual. World Health Organization. 2019. URL: https://apps.who.int/iris/handle/10665/329535 [accessed 2021-01-11]
38. Chopra M, Bhutta Z, Chang Blanc D, Checchi F, Gupta A, Lemango ET, et al. Addressing the persistent inequities in immunization coverage. Bull. World Health Organ 2019 Jan 10;98(2):146-148. [doi: 10.2471/blt.19.241620]
39. Shen A, Fields R, McQuestion M. The future of routine immunization in the developing world: challenges and opportunities. Glob Health Sci Pract 2014 Dec 10;2(4):381-394. [doi: 10.9745/ghsp-d-14-00137]
40. Surveillance standards for vaccine-preventable diseases, 2nd ed. World Health Organization. 2018. URL: https://apps.who.int/iris/handle/10665/275754 [accessed 2021-07-22]
41. Kim SS, Patel M, Hinman A. Use of m-Health in polio eradication and other immunization activities in developing countries. Vaccine 2017 Mar 07;35(10):1373-1379. [doi: 10.1016/j.vaccine.2017.01.058] [Medline: 28190744]
42. Bawa S, Shuaib F, Saidu M, Ningi A, Abdullahi S, Abba B, et al. Conduct of vaccination in hard-to-reach areas to address potential polio reservoir areas, 2014-2015. BMC Public Health 2018 Dec 13;18(Suppl 4):1312-1120 [FREE Full text] [doi: 10.1186/s12889-018-6194-y] [Medline: 30541501]
43. Haskew J, Kenyi V, William J, Alum R, Puri A, Mostafa Y, et al. Use of mobile information technology during planning, implementation and evaluation of a polio campaign in South Sudan. PLoS One 2015;10(8):e0135362 [FREE Full text] [doi: 10.1371/journal.pone.0135362] [Medline: 26252383]
44. Oh DH, Dabbagh A, Goodson JL, Strebel PM, Thapa S, Giri JN, et al. Real-time monitoring of vaccination campaign performance using mobile phones - Nepal, 2016. MMWR Morb Mortal Wkly Rep 2016 Oct 07;65(39):1072-1076 [FREE Full text] [doi: 10.15585/mmwr.mm6539a5] [Medline: 27711034]
45. Teng JE, Thomson DR, Lascher JS, Raymond M, Ivers LC. Using Mobile Health (mHealth) and geospatial mapping technology in a mass campaign for reactive oral cholera vaccination in rural Haiti. PLoS Negl Trop Dis 2014;8(7):e3050 [FREE Full text] [doi: 10.1371/journal.pntd.0003050] [Medline: 25078790]
46. Chandir S, Dharma VK, Siddiqi DA, Khan AJ. Feasibility of using global system for mobile communication (GSM)-based tracking for vaccinators to improve oral poliomyelitis vaccine campaign coverage in rural Pakistan. Vaccine 2017 Sep 05;35(37):5037-5042. [doi: 10.1016/j.vaccine.2017.07.026] [Medline: 28802756]
47. The importance of high-quality AFP surveillance data in the fight to eradicate polio. EHealth Africa - Building Stronger Health Systems in Africa. 2019 Feb 22. URL: https://www.ehealthafrica.org/blog/2019/2/18/the-importance-of-high-quality-afp-surveillance-data-in-the-fight-to-eradicate-polio [accessed 2021-01-18]
48. The use of geospatial data and technologies in support of immunization microplanning in Myanmar Implementation Report (2017 - 2018). United Nations Children's Emergency Fund. 2018. URL: https://www.unicef.org/media/58036/file [accessed 2021-01-12]
49. Guidance on the use of geospatial data and technologies in immunization programs: overview and managerial considerations for in-country strengthening. United Nations Children's Emergency Fund. 2018. URL: https://www.unicef.org/media/58181/file [accessed 2021-01-12]
50. Ebener S, Roth S, Khetrapal S. Building capacity for geo-enabling health information systems: supporting equitable health services and well-being for all (ADB Briefing Paper no 88). Asian Development Bank. 2018. URL: https://www.adb.org/publications/building-capacity-geo-enabling-health-information-systems [accessed 2021-01-12]

51. Leveraging geospatial technologies and data to strengthen immunisation programmes. Gavi, the Vaccine Alliance, in collaboration with UNICEF. 2021 Apr. URL: https://www.gavi.org/sites/default/files/evaluations/Leveraging-Geospatial-Technologies.pdf [accessed 2021-06-04]
52. WHO GIS Centre for Health. World Health Organization. URL: https://www.who.int/data/gis [accessed 2021-06-04]

Abbreviations

GAVI: Gavi, the Vaccine Alliance
RED: Reach Every District
UNICEF: United Nations Children's Emergency Fund
WHO: World Health Organization

Edited by G Eysenbach; submitted 19.04.21; peer-reviewed by E Chukwu, M Das; comments to author 18.05.21; revised version received 14.06.21; accepted 05.07.21; published 03.08.21.

Please cite as:
Chaney SC, Mechael P, Thu NM, Diallo MS, Gachen C
Every Child on the Map: A Theory of Change Framework for Improving Childhood Immunization Coverage and Equity Using Geospatial Data and Technologies
J Med Internet Res 2021;23(8):e29759
URL: https://www.jmir.org/2021/8/e29759
doi:10.2196/29759
PMID:34342584

©Sarah Cunard Chaney, Patricia Mechael, Nay Myo Thu, Mamadou S Diallo, Carine Gachen. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 03.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Viewpoint

Podcasts for the Delivery of Medical Education and Remote Learning

Julliana Newman[1], BA; Andrew Liew[2], PhD; Jon Bowles[3], BA; Kelly Soady[4], PhD; Steven Inglis[2], PhD

[1]Gilead Sciences, Melbourne, Australia
[2]Oxford PharmaGenesis, Melbourne, Australia
[3]Oxford PharmaGenesis, Oxford, United Kingdom
[4]PharmaGenesis London, London, United Kingdom

Corresponding Author:
Julliana Newman, BA
Gilead Sciences
Level 6, 417 St Kilda Road
Melbourne, 3004
Australia
Phone: 61 0438577290
Email: Julie.Newman@gilead.com

Abstract

Podcasts are increasingly being recognized as an effective platform to facilitate the continuous professional development (CPD) of health care professionals (HCPs). Compared with face-to-face meetings and other more traditional forms of CPD, podcasts allow for flexible learning and are less expensive to develop. Podcasts are at the cutting edge of digital education and can be an important element of a pharmaceutical company's multichannel communications plan to improve HCP engagement and CPD in specific therapy areas. However, developing a successful podcast can have significant challenges. In this viewpoint paper, we provide our perspectives on medical podcasts as a medium for educating HCPs in the digital age. We describe our experience in developing an HIV-focused podcast for Australian HCPs, creating a series that has now expanded to other therapy areas in several countries. Practical considerations and unique challenges associated with industry-sponsored podcasts are outlined. Overall, we believe that the process of developing a podcast can be a challenging but rewarding experience, and CPD delivered via podcasting should be more routinely considered by pharmaceutical companies.

(J Med Internet Res 2021;23(8):e29168) doi:10.2196/29168

KEYWORDS

digital; hepatitis C virus; health care professionals; hepatology; HIV; continuous professional development; podcasts; remote learning; virology

Introduction

Podcasts have gained in popularity during the mid-2000s and are now widely available on the internet and on various on-demand channels [1]. Although many podcasts have been developed for entertainment, they are increasingly being used as a platform for medical education [2-5], and recent review articles have shown that a wide variety of medical podcasts covering numerous medical specialties are available [2,6,7]. Medical journals such as *The New England Journal of Medicine* [8] and *The Lancet* [9] are using podcasts to discuss contemporary topics in medicine, further supporting the idea that medical podcasts can supplement more traditional forms of learning, such as peer-reviewed journal articles.

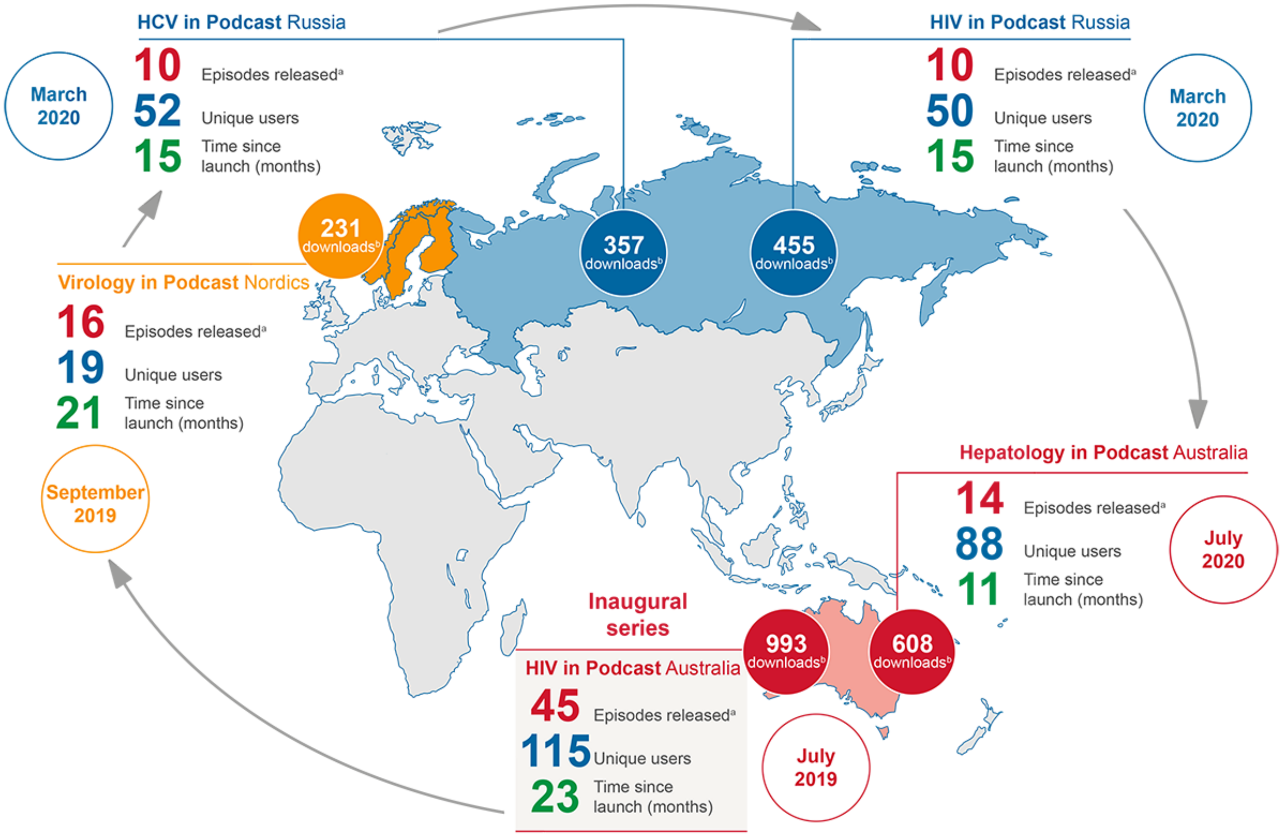
This rise in the popularity of podcasts among the medical community is probably due to several factors. Podcasts allow for rapid dissemination of up-to-date data [10] and provide a range of perspectives to listeners (eg, patient perspectives). The availability of on-demand content helps listeners to learn on the go, which is especially important for busy health care professionals (HCPs) [10]. Unlike face-to-face lectures or congress presentations, podcasts avoid the need for and the expense of travel and give opportunities to HCPs from resource-limited settings to access new perspectives and data. Medical podcasts typically cater to a wide audience with different levels of knowledge (eg, clinicians, medical students, and other HCPs, nurses, pharmacists, and allied HCPs).

Medical podcasts can be grouped into two broad categories: industry-sponsored and general medical podcasts.

Industry-sponsored podcasts are developed by (or in collaboration with) pharmaceutical companies, are typically developed exclusively for the continuing professional development (CPD) of HCPs, and focus on a particular therapy area or product. General medical podcasts are often developed by individuals or medical societies for a wider audience (eg, medical students or the general public, as well as HCPs) and cover a wide range of topics [2].

We noted that there were very few digitally based CPD resources available circa 2018 for Australian HCPs who care for people living with HIV. In response, we launched *HIV in Podcast* Australia, a podcast that provides the latest insights into HIV care from leading experts. *HIV in Podcast* Australia was initially hosted on a web-based platform exclusive to HCPs [11] and in July 2019, we launched a dedicated app to house the podcast. Since then, the series has expanded to include additional therapy areas and regions, and now comprises the following: *HIV in Podcast* Australia, *Hepatology in Podcast* Australia, *Virology in Podcast* Nordics, *HCV in Podcast* Russia, and *HIV in Podcast* Russia (Figure 1). As appropriate, content was tailored to region-specific health challenges. Apart from the 2 Russian series, all podcasts were recorded in English.

Figure 1. Overview of the *in Podcast* series and podcast metrics provided by the hosting platform. [a]As of May 31, 2021. [b]Audio file download to app or live stream across series.


HCV in Podcast Russia
10 Episodes released[a]
52 Unique users
15 Time since launch (months)
March 2020
HIV in Podcast Russia
10 Episodes released[a]
50 Unique users
15 Time since launch (months)
March 2020
231 downloads[b]
357 downloads[b]
455 downloads[b]
Virology in Podcast Nordics
16 Episodes released[a]
19 Unique users
21 Time since launch (months)
September 2019
Hepatology in Podcast Australia
14 Episodes released[a]
88 Unique users
11 Time since launch (months)
July 2020
Inaugural series
HIV in Podcast Australia
45 Episodes released[a]
115 Unique users
23 Time since launch (months)
993 downloads[b]
608 downloads[b]
July 2019

In this paper, based on our own experience with the *in Podcast* program, we describe practical considerations and challenges associated with developing, distributing, and measuring the success of industry-sponsored podcasts. Several guides on creating effective general medical podcasts have been published [12-15]. However, to the best of our knowledge, there are no publications that discuss the unique challenges associated with developing industry-sponsored podcasts. Overcoming some of these challenges could contribute to an increase in the use of this CPD tool for HCPs.

Practical Considerations for Industry-Sponsored Podcasts

To develop high-quality and engaging medical podcasts, a few basic elements are implicit. In short, these include the recruitment of a suitably qualified host with medical training and therapy area experience, ideally with medical journalism experience. Equally, interviewees should be regional or international experts in their field and be able to discuss contemporary topics concisely. Podcasts should be "bite-sized," with recommended lengths of 10-20 minutes [3], and recorded and produced by professional sound engineers.

In our experience, development of industry-sponsored podcasts often present other unique challenges and considerations, which we describe in further detail below.

Podcast Hosting Platforms

Podcast episodes need to be hosted on an appropriate digital platform for access by listeners [12]. A myriad of hosting options are available (eg, custom-built websites or apps and paid and free platforms) [12], and choosing a platform will depend on budgets, compliance requirements, and timelines. An ideal hosting platform should be secure and easy to use, provide reliable and standardized metrics, and be adaptable for local and regional regulatory requirements regarding distribution of content to HCPs. In general, access to podcast episodes can be achieved via the webpage of the hosting platform or a dedicated mobile app. Such apps should be easy to download, have intuitive interfaces, be compatible with both Apple and Android operating systems, and allow download and storage of episodes for offline listening. Based on these considerations, we chose a specialist podcast hosting platform [16] for the *in Podcast* series that is free of charge to the end user, with an optional associated app to provide HCPs with the convenience of accessing content via their mobile device. The app is available in multiple languages (to mirror the podcasts). It is important to note that developing such an app requires significant investment, likely including partnering with companies or individuals with app development expertise.

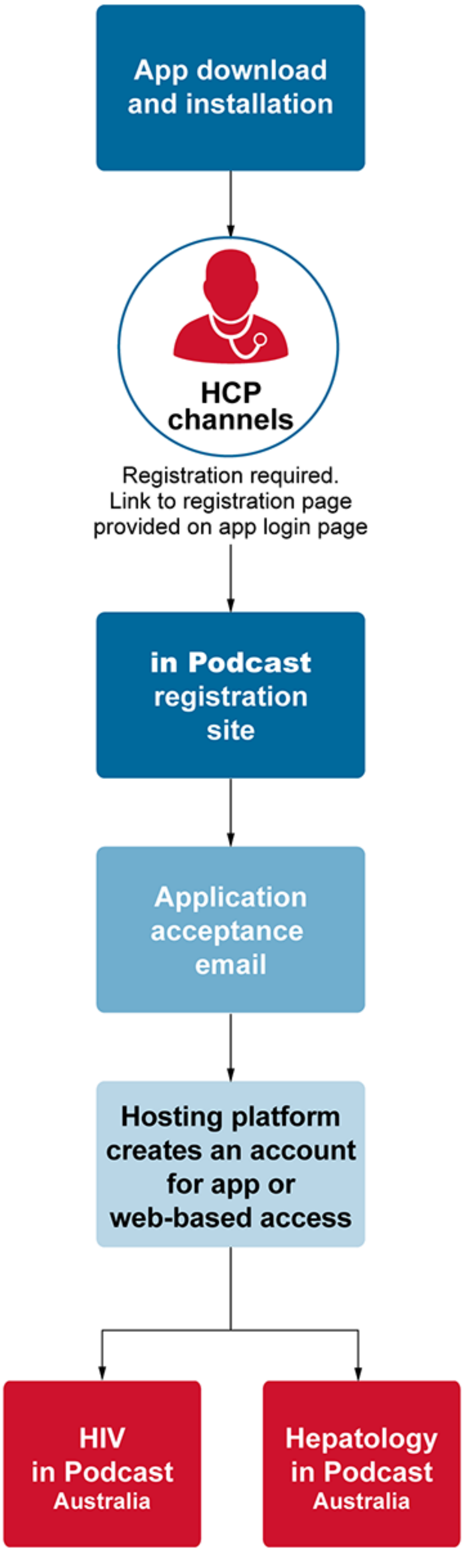
Regulatory Considerations

A challenge encountered with industry-sponsored podcasts is ensuring that the discussions related to medications or clinical practice are in line with the approved indication for that region, fair, balanced, and consistent with local and regional guidelines governing communication with HCPs. For example, guests might describe treatments that are not approved in the country where the podcast is released, or they might discuss data from studies of unlicensed uses for approved agents; thus, content could be deemed promotional or noncompliant with local guidelines [17,18]. Medical writers, together with regulatory personnel of the sponsor company, play an important role in managing content development. In the *in Podcast* program, the medical writing team were involved in all aspects of episode development, including podcast recording sessions, to ensure that discussions between the podcast host and guests were accurate and reflected the approved indications for any treatment discussed ("on label"). The medical writing team were also responsible for postrecording activities such as transcribing the recording and ensuring that all scientific claims were supported by current evidence from peer-reviewed journal papers or government guidelines, with a reference list available for every podcast episode. We believe that medical writers play a crucial role, in partnership with the regulatory team, in ensuring that all podcast content is fair and balanced.

The *in Podcast* series has been launched across different regions; therefore, we had to ensure compliance with regulatory and privacy requirements for each of the relevant countries. The introduction of the European Union General Data Protection Regulation (GDPR) in 2018 implied ensuring the protection of personal data, privacy, and consent of podcast users. To achieve this, we engaged medical, regulatory, business compliance, and legal stakeholders in the region and therapy area when planning each series. Some hosting platforms collect personal data (eg, email addresses) for metrics analysis, so it is important to utilize a GDPR-compliant hosting platform.

In some countries, discussion or promotion of prescription medicines is only permitted among HCPs. As such, industry-sponsored podcast topics may only be appropriate for an HCP audience (eg, content that could be perceived as promotion of a pharmaceutical product to a non-HCP audience). To ensure that only HCPs could access the content, we created a registration webpage that allowed us to verify a user's HCP status via their medical/professional registration numbers, or enabled the HCP to self-verify their status, depending on local regulatory requirements. Once a user's HCP status is verified, they are sent detailed instructions on how to download the podcasting app, create a personal podcast account, and access episodes via the app or a web browser. A summary of the user registration process we used is shown in Figure 2. Because HCPs register through a country-specific registration page, they are given access to content available in their region only. This system enables 1 platform and mobile app to be used across multiple regions.

Figure 2. Podcast registration flow for Australian HCPs accessing *HIV in Podcast* Australia or *Hepatology in Podcast* Australia episodes. Channel access is determined by geographical location upon registration, so that Australian HCPs will only have access to Australian channels. HCP: health care professional.


App download and installation
HCP channels
Registration required. Link to registration page provided on app login page
in Podcast registration site
Application acceptance email
Hosting platform creates an account for app or web-based access
HIV in Podcast Australia
Hepatology in Podcast Australia

Accreditation of Podcasts

The credibility of industry-sponsored podcasts can be enhanced through accreditation by a professional body (a medical society or college), thereby allowing HCPs to use the podcasts for CPD. Medical practitioners registered with the Medical Board of Australia, for example, need to complete a minimum number of CPD hours that are relevant to their practice [19]. Accreditation of podcasts may thus help to increase their uptake by the target audience. To gain accreditation, podcast transcripts and supporting references for each episode of *HIV in Podcast* Australia were reviewed by the Australasian Society of HIV Medicine [20]. Users were encouraged to self-report their listening activity to Australasian Society of HIV Medicine to claim CPD points and maintain their registration as prescribers of HIV treatment. It is important to note that the role of CPD points in continuing professional medical education and accreditation requirements varies across regions. The medical writing team plays an important role in ensuring that all podcast materials meet the accreditation standards of the relevant accrediting society.

Audience Recruitment

Recruiting an appropriate and sizable audience can be challenging. Promoting general medical podcasts typically involves the use of social media platforms (eg, blogs, Facebook, and Twitter) [13] or joining a podcast network [21]. However, if applied to industry-sponsored podcasts, this approach may be considered an untargeted promotional activity and tends to be avoided by pharmaceutical sponsors. Instead, we employed several focused strategies to increase the awareness of the podcasts among HCPs, including the use of hard copy and electronic mailouts to HCPs on subscriber lists provided by industry partners, publishing printed advertisements in peer-reviewed journals or scientific magazines, and conducting promotional activities at scientific meetings (setting up booths, handing out flyers, and conducting in-person registrations). We found that electronic mailouts reach a wider, more targeted audience than the other methods but may require the use of external platforms (eg, Mailchimp) or third parties (eg, advertising or event management agencies who have access to lists of potential subscribers). We found that an advantage of promoting podcasts at scientific meetings was that it provided us with opportunities to gain valuable feedback on the podcasts. We recommend that multichannel promotional strategies be used, if possible, to maximize podcast uptake.

Measuring the "Success" of Podcasts

Regularly measuring the uptake of podcasts by HCPs is essential for evaluating the performance of a podcast and can be an indicator of the success of the strategies used to attract listeners. The number of downloads is one of the most frequently used digital performance metrics [3], although other metrics are available (eg, the number of podcast users) [22]. At the time of writing (May 31, 2021), *HIV in Podcast* Australia had achieved the largest number of total downloads and unique users (993 downloads; 115 unique users) followed by *Hepatology in Podcast* Australia (608 downloads; 88 unique users; Figure 1). In Europe and Russia, *HIV in Podcast* Russia achieved the largest number of total downloads (455 downloads), whereas *HCV in Podcast* Russia had the highest number of unique users (52 users). Relative to the small number of unique users, the large number of downloads suggests that users likely downloaded multiple podcast episodes, which indicates continued engagement with content over time. In our experience, the number of downloads can be influenced by various factors including therapy area, size of the target audience, geographical region, and how readily the target audience can access the podcast. This makes direct comparisons of podcast performance difficult. For example, average monthly downloads for the *in Podcast* series, housed on our app in which HCPs must register to gain access, were between 56 downloads per month for *Hepatology in Podcast* Australia and 11 downloads per month for *Virology in Podcast* Nordics (Figure 1). In contrast, publicly available medical podcasts have been reported to achieve between 217 and 10,000 downloads per month [10,23,24]. Comparisons with other industry-sponsored podcasts are challenging owing to a lack of published data.

Another important consideration is that download metrics provided by GDPR-compliant hosting platforms contain anonymized user identifications. This makes it difficult to profile podcast users by HCP type (eg, proportion of physicians vs nurses), which may be of specific interest to the sponsor. Additional qualitative measures of success may therefore add value. For example, short surveys can be included in electronic mailouts or conducted via face-to-face interviews with the target audience [10]. These qualitative data can be used to understand if, and the extent to which, listeners are enjoying the series and whether it may be changing their clinical practice. Additionally, podcast performance metrics should be codeveloped with industry partners and realistic goals set for measuring "success" (eg, a prespecified proportion of licensed prescribers within a therapy area in 1 country). We believe that combining both quantitative and qualitative metrics provides more holistic insights into the overall performance of the podcast.

Viewpoint

There is a growing interest from the pharmaceutical industry to invest in digital innovation, such as podcasts, to increase awareness of diseases, clinical practice updates, and available therapies. At present, most medical podcasts have been created by individuals and academic societies rather than pharmaceutical companies. Medical podcasts are an important medium for delivering specific and targeted CPD to HCPs and could be considered more routinely by the pharmaceutical industry as part of multichannel communication plans to improve HCP engagement and education. However, very little published information is available on the considerations and challenges associated with creating industry-sponsored podcasts.

In our opinion, one key challenge is the limitation around granular metrics data on industry-sponsored podcasts, making it difficult to demonstrate a return on investment. Close collaboration with industry partners will be needed to define the target audience and to help ensure that meaningful metrics are collected. The inclusion of qualitative measures (eg, user surveys) is also recommended to determine if podcast uptake is associated with changes in clinical practice.

The requirement for users to register and be verified as HCPs before being able to access the podcasts may also be a significant "digital barrier" to the uptake of industry-sponsored podcasts. We suspect that multistep registration processes (Figure 2) and the need to create a personal podcast account may act as a deterrent for busy HCPs to quickly access podcast content. However, these are necessary steps from an industry compliance perspective in many jurisdictions. Possible options for overcoming this challenge could include setting up preregistrations and involving medical science liaison officers of companies to help users with the registration process. If regulations allow, creating podcast apps that support single-click access from an invitation email for HCPs who have been preregistered and verified could reduce the steps required for authentication and prevent possible "password fatigue," which may result from users needing to manage multiple passwords and accounts.

Overcoming these barriers could help increase the use of this flexible and versatile educational platform among the medical community. Indeed, recent innovations have highlighted that podcasting will be an important medium for educating HCPs in the future. For example, several Adis-affiliated journals allow the publication of peer-reviewed podcast articles on clinically-relevant topics (eg, conference data, treatment innovations, and expert opinions on drugs and disease therapies) [25]. We believe that the challenges and potential solutions we have outlined are not only relevant for podcasting but also broadly applicable to the various forms of industry-sponsored digital education (eg, webinars and education modules) that are increasingly being used.

Conclusions

Podcasting is a novel method for delivering medical education and can be a suitable alternative to traditional face-to-face and print-based education and incorporated into CPD programs. Careful planning is needed to ensure the success of industry-sponsored podcasts. With further uptake and optimization of this form of medical education, industry-sponsored podcasts may play an important role in shaping change in clinical practice and improving patient care in the future.

Acknowledgments

We thank the members of the *in Podcast* team at Oxford PharmaGenesis (Vanessa Bateman, Lucy Eberhard, Robert Goodwin, Cara Kingston, Svetha Sankar, Tanya Stezhka, and Vanessa Wielgosz) and Gilead Sciences (Hassan Aladdin, Mike Elliott, Damien Fagan, Aurelie Freismuth, Janniche Hammer, Geraldine Reilly, Natalia Tikhonova, and Tatiana Usacheva) for their contributions to the series. We also thank the speakers and series hosts for contributing to the podcasts, especially Dr Brad McKay at East Sydney Doctors (Sydney, New South Wales, Australia), for playing a key role in the development and success of the inaugural *HIV in Podcast* Australia series. We are also grateful to Svetha Sankar for critical review of this manuscript.

Conflicts of Interest

JN is an employee of Gilead Sciences. JB, SI, AL, and KS are employees of Oxford PharmaGenesis, which has received funding to support development of the *in Podcast* series and this article, from Gilead Sciences.

References

1. The Infinite Dial 2018. Edison Research. 2018. URL: http://www.edisonresearch.com/wp-content/uploads/2018/03/Infinite-Dial-2018.pdf [accessed 2021-03-18]
2. Little A, Hampton Z, Gronowski T, Meyer C, Kalnow A. Podcasting in medicine: a review of the current content by specialty. Cureus 2020 Jan 21;12(1):e6726 [FREE Full text] [doi: 10.7759/cureus.6726] [Medline: 32104642]
3. Cho D, Cosimini M, Espinoza J. Podcasting in medical education: a review of the literature. Korean J Med Educ 2017 Nov 29;29(4):229-239 [FREE Full text] [doi: 10.3946/kjme.2017.69] [Medline: 29207454]
4. Back D, von Malotky J, Sostmann K, Hube R, Peters H, Hoff E. Superior gain in knowledge by podcasts versus text-based learning in teaching orthopedics: a randomized controlled trial. J Surg Educ 2017;74(1):154-160. [doi: 10.1016/j.jsurg.2016.07.008] [Medline: 27651055]
5. Rodman A, Trivedi S. Podcasting: a roadmap to the future of medical education. Semin Nephrol 2020 May;40(3):279-283. [doi: 10.1016/j.semnephrol.2020.04.006] [Medline: 32560776]
6. Chen Z, Melon J. Evolution of social media: review of the role of podcasts in gynaecology. Int Urogynecol J 2018 Apr;29(4):477-480. [doi: 10.1007/s00192-017-3523-0] [Medline: 29188323]
7. O'Connor S, Daly C, MacArthur J, Borglin G, Booth R. Podcasting in nursing and midwifery education: an integrative review. Nurse Educ Pract 2020 Aug;47:102827 [FREE Full text] [doi: 10.1016/j.nepr.2020.102827] [Medline: 32763834]
8. NEJM Audio Interview. New England Journal of Medicine. 2020. URL: https://www.nejm.org/multimedia/audio-interview [accessed 2021-03-18]
9. The Lancet Voice. The Lancet. 2020. URL: https://www.thelancet.com/the-lancet-voice [accessed 2021-03-18]
10. Malecki S, Quinn K, Zilbert N, Razak F, Ginsburg S, Verma A, et al. Understanding the use and perceived impact of a medical podcast: qualitative study. JMIR Med Educ 2019 Sep;5(2):e12901 [FREE Full text] [doi: 10.2196/12901] [Medline: 31538949]
11. Praxhub. Praxhub. 2018. URL: https://praxhub.com [accessed 2021-03-18]
12. Hurst E. Getting started with podcasting. J Hosp Librariansh 2019 Jul;19(3):277-283. [doi: 10.1080/15323269.2019.1628575]
13. Ahn J, Inboriboon P, Bond M. Podcasts: accessing, choosing, creating, and disseminating content. J Grad Med Educ 2016 Jul;8(3):435-436 [FREE Full text] [doi: 10.4300/JGME-D-16-00205.1] [Medline: 27413451]
14. Sandars J. Twelve tips for using podcasts in medical education. Med Teach 2009 May;31(5):387-389. [doi: 10.1080/01421590802227958] [Medline: 19140038]

15. Lichtenheld A, Nomura M, Chapin N, Burgess T, Kornegay J. Development and implementation of an emergency medicine podcast for medical students: EMIGcast. West J Emerg Med 2015 Nov;16(6):877-878 [FREE Full text] [doi: 10.5811/westjem.2015.9.27293] [Medline: 26594282]
16. PodBean. PodBean. 2021. URL: https://www.podbean.com [accessed 2021-06-11]
17. Code of Practice for the Pharmaceutical Industry 2019. Association of the British Pharmaceutical Industry. 2019. URL: https://www.abpi.org.uk/media/6655/abpi-code-of-practice-2019.pdf [accessed 2021-03-18]
18. Code of Conduct: edition 19. Medicines Australia. 2020. URL: https://www.medicinesaustralia.com.au/wp-content/uploads/2020/11/20200108-PUB-Edition-19-FINAL.pdf [accessed 2021-03-18]
19. Codes, guidelines and policies. Medical Board of Australia. 2020. URL: https://www.medicalboard.gov.au/Codes-Guidelines-Policies/FAQ/FAQ-general-registration-CPD.aspx [accessed 2021-03-18]
20. ASHM. Australasian Society for HIV, Viral Hepatitis and Sexual Health Medicine. 2021. URL: https://www.ashm.org.au [accessed 2021-06-11]
21. Hurst E. Podcasting in medical education and health care. J Hosp Librariansh 2019 Jul;19(3):214-226. [doi: 10.1080/15323269.2019.1628564]
22. PodBean statistics overview. PodBean. 2020. URL: https://help.podbean.com/support/solutions/articles/25000018937-podcast-statistics-overview [accessed 2021-03-18]
23. Patrick M, Stukus D, Nuss K. Using podcasts to deliver pediatric educational content: development and reach of PediaCast CME. Digit Health 2019 Feb;5:2055207619834842 [FREE Full text] [doi: 10.1177/2055207619834842] [Medline: 30834137]
24. Nwosu A, Monnery D, Reid V, Chapman L. Use of podcast technology to facilitate education, communication and dissemination in palliative care: the development of the AmiPal podcast. BMJ Support Palliat Care 2017 Jun;7(2):212-217 [FREE Full text] [doi: 10.1136/bmjspcare-2016-001140] [Medline: 27580942]
25. Adis digital features. Springer Healthcare. 2021. URL: https://springerhealthcare.com/expertise/publishing-digital-features [accessed 2021-05-26]

Abbreviations

CPD: continuing professional development
GDPR: General Data Protection Regulation
HCP: health care professional

Edited by R Kukafka; submitted 07.04.21; peer-reviewed by J Espinoza, R Lakhtakia; comments to author 16.05.21; revised version received 30.06.21; accepted 27.07.21; published 27.08.21.

Please cite as:
Newman J, Liew A, Bowles J, Soady K, Inglis S
Podcasts for the Delivery of Medical Education and Remote Learning
J Med Internet Res 2021;23(8):e29168
URL: https://www.jmir.org/2021/8/e29168
doi:10.2196/29168
PMID:34448719

©Julliana Newman, Andrew Liew, Jon Bowles, Kelly Soady, Steven Inglis. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Web-Based Guidance Through Assisted Reproductive Technology (myFertiCare): Patient-Centered App Development and Qualitative Evaluation

Ellen Marie Sparidaens[1], MD; Rosella P M Hermens[2], PhD; Didi D M Braat[1], MD, PhD; Willianne L D M Nelen[1], MD, PhD; Kathrin Fleischer[3], MD, PhD

[1]Department of Obstetrics and Gynaecology, Radboud University Medical Center, Nijmegen, Netherlands
[2]Scientific Institute for Quality of Healthcare (IQ Healthcare), Radboud University Medical Center, Nijmegen, Netherlands
[3]The Fertility Partnership Center of Reproductive Medicine, Düsseldorf, Germany

Corresponding Author:
Ellen Marie Sparidaens, MD
Department of Obstetrics and Gynaecology
Radboud University Medical Center
PO Box 9101
Nijmegen, 6500HB
Netherlands
Phone: 31 (0)243617351
Email: ellenmarie.sparidaens@radboudumc.nl

Abstract

Background: Providing patient-centered fertility care is known to improve quality of life and can reduce anxiety and depression. In a previous study, we established the need for a web-based app providing personalized information and interactive functionalities among couples undergoing intracytoplasmic sperm injection with surgically retrieved sperm.

Objective: This study aimed to design, develop, and qualitatively evaluate a multifaceted web-based app for infertile couples undergoing intracytoplasmic sperm injection with surgically retrieved sperm during their treatment trajectory.

Methods: The web-based app was developed in three phases: (1) we established a patient-centered functional design, (2) developed the app in collaboration with medical and technical professionals, and (3) qualitatively evaluated the app among couples using a think-aloud method.

Results: The basis of the app is the couple’s visualized treatment trajectory. The app provides personalized and interactive functionalities; for example, customized information and communication options. During qualitative evaluation, myFertiCare was highly appreciated and received a median score of 8 out of 10. The main improvements made upon conclusion of the think-aloud sessions were related to faster login and easier app navigation.

Conclusions: A patient-centered web-based app aimed at guiding couples through their fertility treatment course was systematically designed, developed, and positively evaluated by patients and medical and technical professionals.

(J Med Internet Res 2021;23(8):e25389) doi:10.2196/25389

KEYWORDS

eHealth; infertility; interactive; mobile apps; patient education; patient-centered care; personalized; topic

Introduction

According to the Institute of Medicine, health care should be safe, effective, timely, efficient, equitable, and patient-centered [1]. Patient-centeredness is described as “providing care respectful of and responsive to individual patient preferences, needs and values, and ensuring that patient values guide all clinical decisions” [1]. Providing patient-centered fertility care can improve quality of life and can reduce anxiety and depression [2]. It has been reported that patient-centeredness in fertility care needs improvement [3,4]. Crucial aspects of patient-centeredness in fertility care are provision of adequate information, continuity of care, and active involvement of patients in their treatment course [5-8]. The internet is known to be an important source of information and support for infertile couples [9-11]. A possible instrument for improving patient-centeredness is the use of eHealth tools; that is, web-based apps in health care [12]. Previous eHealth initiatives in fertility care aimed mainly at information provision, support, and mental health promotion [13]. These initiatives only contained few interactive web-based components. Aarts et al [13] concluded that these initiatives could be improved by including more interactive and dynamic elements as their key components. Infertile couples are known to specifically prefer personalized information and appreciate being able to communicate with both their treatment team and fellow patients [3,10]. eHealth tools are a promising strategy to empower patients in managing their own treatment trajectory.

In a previous study [14], we established the need for a web-based app providing personalized information and interactive functionalities among couples undergoing intracytoplasmic sperm injection (ICSI) with surgically retrieved sperm. We hypothesized that a web-based app is specifically suitable for this group of patients because of the psychological and physical burden of the multidisciplinary treatment on both partners. Therefore, the aim of this study was to design, develop, and qualitatively evaluate a multifaceted web-based app for use by couples undergoing ICSI with surgically retrieved sperm during their treatment trajectory.

Methods

Systematic Approach

myFertiCare was developed in three phases: (1) we established a patient-centered, functional design for the app; (2) developed myFertiCare in collaboration with medical and technical professionals; and (3) had myFertiCare qualitatively evaluated for usability, with a think-aloud method.

Phase 1: Establishment of a Patient-Centered Functional Design

The functional design of myFertiCare is based on (1) literature review; (2) interviews with an expert panel comprising a gynecologist, a urologist, an embryologist, an expert in patient-centered innovation, and a board member of Freya, the Dutch association for infertility problems; and (3) interviews with a patient panel. This was part of our previous study [14] on the informational needs of couples undergoing ICSI with surgical sperm retrieval. The patient panel consisted of 11 couples, a number that was determined through data saturation. We conducted semistructured interviews with each couple individually. The data were analyzed using a constant comparative method. The functional design that followed from this process was verified by the clinic's fertility treatment team and supplemented with their suggestions. Both the expert and the patient panels were enthusiastic about the idea of a web-based app to guide couples through their treatment trajectory. The overall opinion was that the more functionalities an app provides, the better the app, so that people are motivated to use it. The participants specifically valued personalized and interactive functionalities. Various functionalities were suggested, such as being able to view appointments, test results, and information about lifestyle advice; information about the clinic's fertility treatment team; and communication with physicians and peers. The interviewees also highlighted the safeguarding of confidential information, which they stated should be at the core of app development [14].

Phase 2: Development of myFertiCare

Based on the preferences of the expert and the patient panels, myFertiCare was developed in close collaboration with medical professionals from the department of Reproductive Medicine, Radboud University Medical Center (RUMC), and technical experts from a Dutch company specializing in eHealth. Together, they formed the project team. Development was an iterative process. The desired functionalities of myFertiCare were categorized by the medical professionals as must-have functionalities that had to be available before the app could be made available on the internet, or as nice-to-have functionalities that could be developed later. Subsequently, the must-have functionalities were developed by the technical experts and tested in a test environment by both the technical experts and the medical professionals. Technical adjustments were made as necessary, and the testing cycle was started over again. Once all the must-have functionalities were developed and tested by the technical and medical experts, myFertiCare was made available through the hospital website, the App Store, and Google Play Store to couples undergoing ICSI with surgical sperm retrieval. myFertiCare was also incorporated in the existing hospital information systems. After the app was live, all the nice-to-have functionalities were developed through the same development cycle. After each functionality was iteratively developed and tested, it was immediately made available to all myFertiCare users. The duration of the whole development trajectory was approximately 1.5 years.

Phase 3: Qualitative Evaluation of myFertiCare for Usability Through the Think-Aloud Method

After all the must-have and nice-to-have functionalities were implemented, we began the qualitative evaluation of myFertiCare. In total, 21 couples, who consecutively visited the fertility clinic, were invited by their physicians to participate in the think-aloud sessions. Six couples agreed to participate, which accounted for 9 participants (4 men and 5 women) (Tables 1 and 2). Reasons for nonparticipation were being too busy, already having too much going on, or simply not wanting to participate. Think-aloud is a research method in which participants verbalize their thought processes while interacting with a tool [15]. It provides a valid source of data about participants' thought processes and can be used effectively in qualitative studies [15]. Our aim was to identify usability flaws and to provide suggestions for design modification.

The participants were individually provided with 16 tasks to perform using myFertiCare while thinking out loud (Figure 1). Of these tasks, 11 were the same for every participant and 5 focused on the specific phase of treatment that an individual was in. By completing these tasks, the participants explored all the functionalities of myFertiCare. The researcher observed the participants and asked questions for clarification where needed. The researcher took notes, also focusing on nonverbal communication of participants. In addition, the sessions were audio-taped. After completing each task, the participants answered 3 task-linked questions ("I find this task easy," "I find this information useful," and "I find this information is in a logical spot") with a 5-point Likert scale to rate responses. They could also add free comments.

Immediately after each think-aloud session, the participants completed a self-developed questionnaire on their experiences using myFertiCare (Figure 2). The questions were about participants' attitudes toward usability, privacy, understandability of information, and the usefulness of various functionalities of myFertiCare. The questionnaire consisted of 20 questions: 17 with responses rated on a 5-point Likert scale, 2 to be answered with "yes" or "no," and 1 open question. Again, the participants could also write free comments.

Table 1. Demographic characteristics of the study participants (N=9).

Characteristic	Men (n=4)	Women (n=5)
Age (years), median (range)	33 (27-47)	30 (28-36)
Daily internet usage in a private setting (minutes), median (range)	60 (45-60)	90 (60-180)
Treatment-related use of the internet, n	3	5
Use of myFertiCare prior to the think-aloud session, n	1	2
Educational status,[a] n		
Low	0	0
Medium	1	2
High	3	3
Ethnic background, n		
Caucasian	3	4
Non-Caucasian[b]	1	1
Those who already have children, n		
Yes	0	0
No	4	5

[a] Low educational status includes no education and lower general secondary education. Medium educational status includes higher general secondary education and intermediate vocational education. High educational status includes higher vocational education and a university degree.

[b] One man from Indonesia and 1 woman from Suriname.

Table 2. Demographic characteristics of the couples who participated in the study (N=6).

Characteristic	Couples
Socioeconomic status,[a] n	
Low	3
Medium	3
High	0
Stage of treatment, n	
Before surgical sperm retrieval	1
After surgical sperm retrieval and before ICSI[b]	2
During first ICSI cycle	2
After at least one full ICSI cycle	1
Duration of infertility (months), median (range)	28 (16-47)

[a] According to the Dutch Social and Cultural Planning Office: Low socioeconomic status included a status score of ≤–1; medium socioeconomic status included a status score between –1 and +1; high socioeconomic status included a status score of >1.

[b] ICSI: intracytoplasmic sperm injection.

Figure 1. Results obtained from think-aloud sessions.

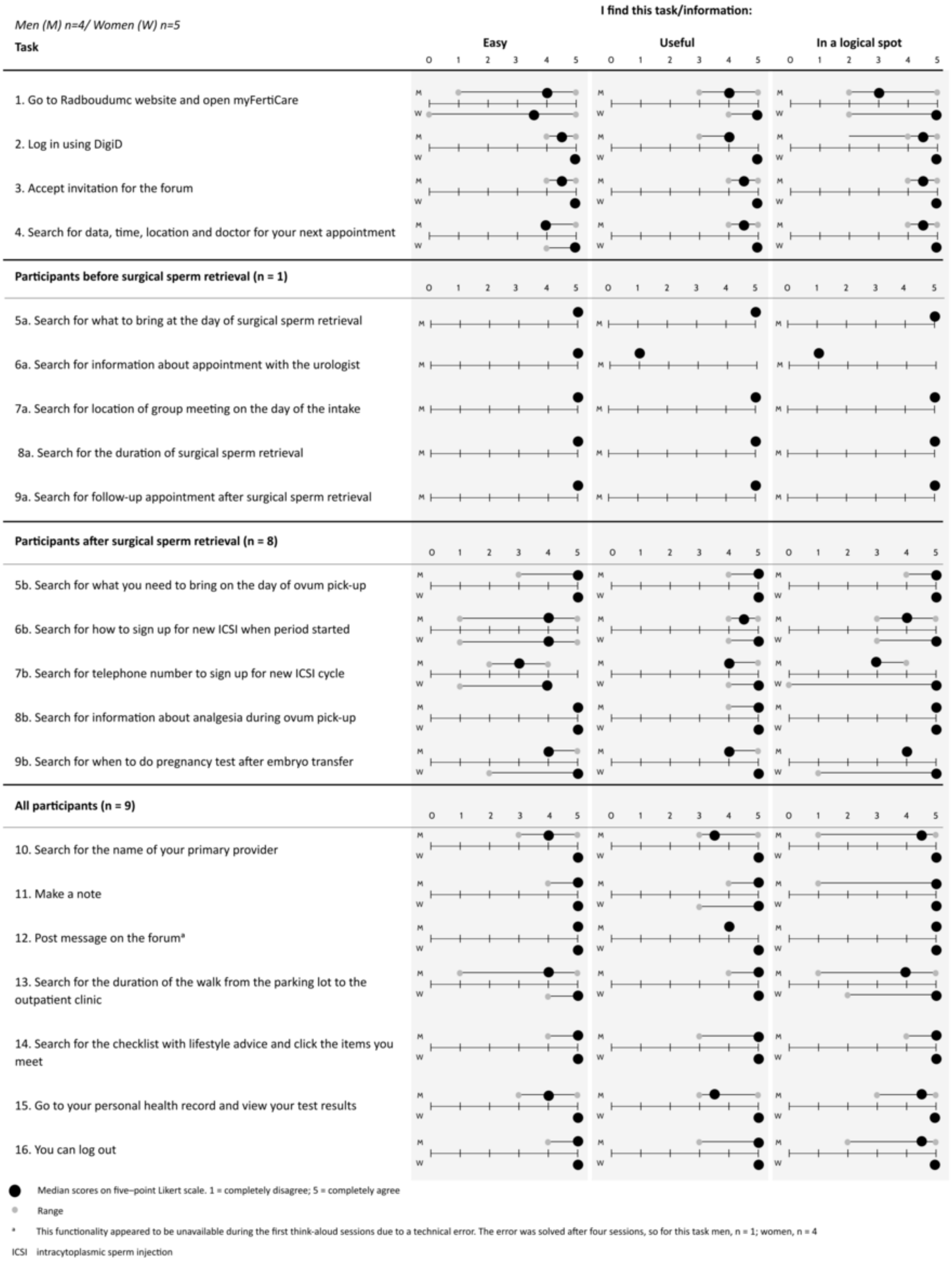
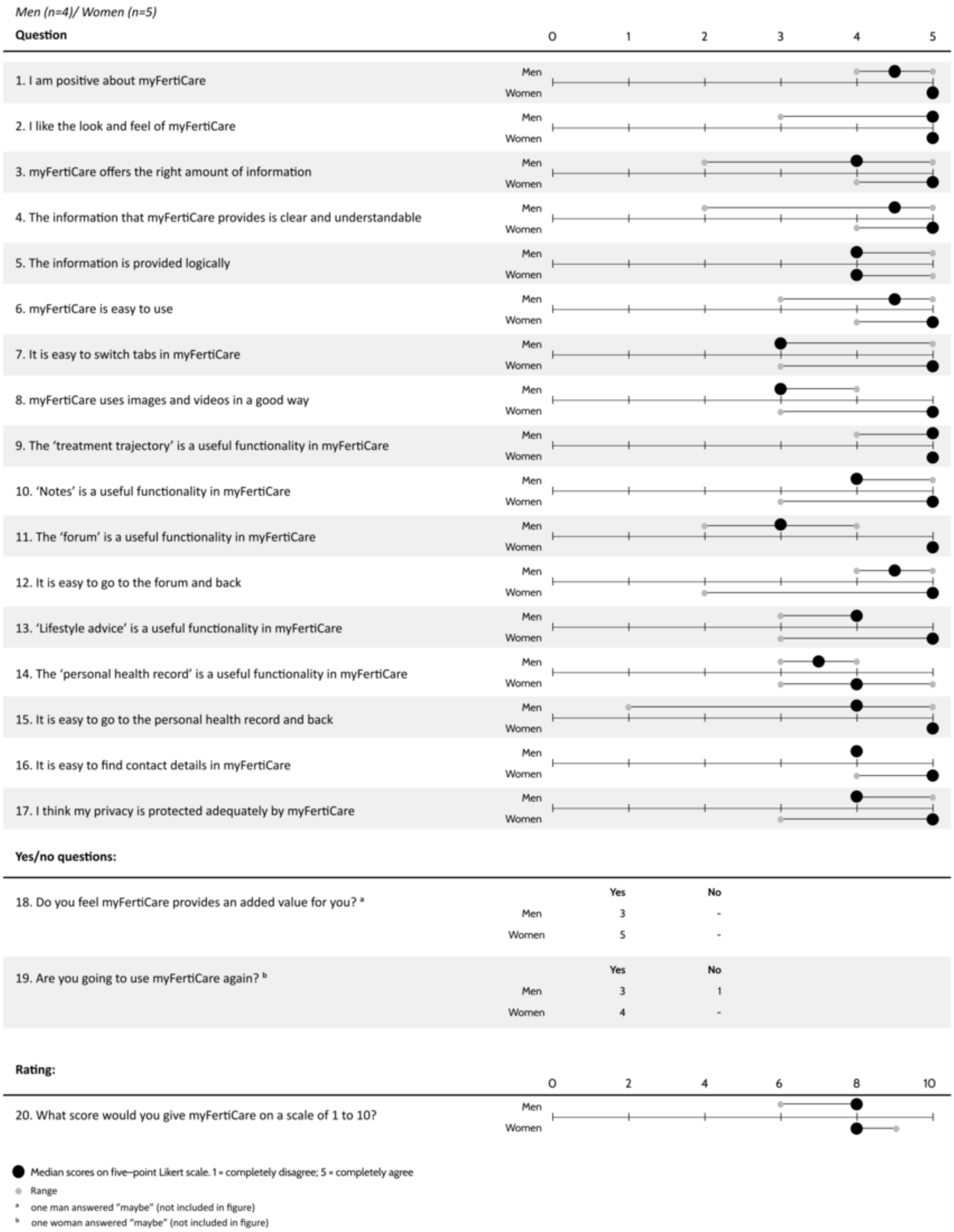


Figure 2. Results obtained from the questionnaire on users' experience with myFertiCare.



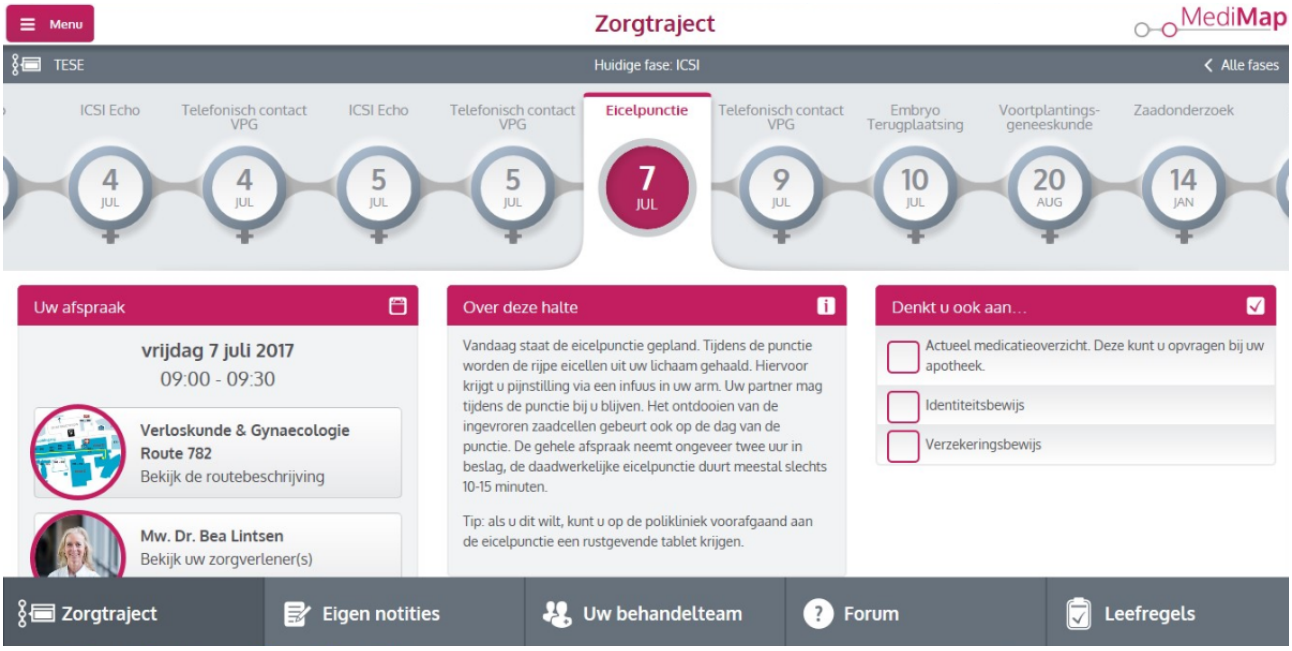
The think-aloud sessions were conducted until saturation was reached. The duration of a session was approximately 15 to 20 minutes. The audio-taped sessions were transcribed verbatim. Data were analyzed anonymously. An open coding method was applied. We coded quotes that identified usability flaws or provided suggestions for modification of the app's design. A second researcher verified the coding process. Differences were discussed until consensus was reached. Ethical approval was proposed but was not required according to the local research ethics committee (CMO Arnhem Nijmegen, file# 2016-2485). All participants provided written informed consent.

Results

Phases 1 and 2: Design and Development of myFertiCare

Based on the results of phase 1 [14], myFertiCare was developed as a web-based app available on the RUMC website, Google Play Store, and Apple App Store for couples undergoing ICSI with surgical sperm retrieval at the RUMC. Patients log in using their digital identity (DigiD), which is provided by the government of the Netherlands to assure safe access to all governmental institutions. This guarantees the safety of couples' medical data. The apps of both partners are synchronized, so that individuals can also see their spouse's information. myFertiCare is free for use and is offered in addition to usual care. A screenshot of the app is provided in Figure 3.

Figure 3. Screenshot of the myFertiCare app.



vrijdag 7 juli 2017
09:00 - 09:30

Verloskunde & Gynaecologie
Route 782
Bekijk de routebeschrijving

Mw. Dr. Bea Lintsen
Bekijk uw zorgverlener(s)

Vandaag staat de eicelpunctie gepland. Tijdens de punctie worden de rijpe eicellen uit uw lichaam gehaald. Hiervoor krijgt u pijnstilling via een infuus in uw arm. Uw partner mag tijdens de punctie bij u blijven. Het ontdooien van de ingevroren zaadcellen gebeurt ook op de dag van de punctie. De gehele afspraak neemt ongeveer twee uur in beslag, de daadwerkelijke eicelpunctie duurt meestal slechts 10-15 minuten.

Tip: als u dit wilt, kunt u op de polikliniek voorafgaand aan de eicelpunctie een rustgevende tablet krijgen.

☐ Actueel medicatieoverzicht. Deze kunt u opvragen bij uw apotheek.
☐ Identiteitsbewijs
☐ Verzekeringsbewijs

myFertiCare contains personalized and interactive functionalities that are divided over 5 tabs:

1. Treatment trajectory: this is the basis of the app. The treatment trajectory is visualized as a subway map in which every stop stands for one of the appointments a couple must have in order to move forward. Couples can see their past appointments and future scheduled appointments with the corresponding data, time, physician, and location, but they can also see future appointments that are not scheduled yet. Thus, couples are better prepared for the upcoming trajectory and know what to expect. Each stop on the subway map provides information about the specific phase of the treatment trajectory and provides advice on how to prepare for the appointment, and, if applicable, informs them of anything they should bring with them for the appointment. Furthermore, users receive support messages before or after certain appointments to comfort them or provide some advice. These support messages are sent via the app or via text message.
2. Notes: users can write notes that are synchronized with their spouse's notes. For example, couples can compose a topic list with questions they want to ask during their upcoming physician's appointment.
3. Care providers: an overview of the whole treatment team is provided through photographs, with an individual's primary care provider on top. Users can ask medical questions to the treatment team, and they are answered within 24 hours.
4. Forum: patients can communicate with peers on the forum. The forum is supervised by a clinician.
5. Lifestyle advice: this is provided as separate checklists for men and women. The aim is to improve treatment outcomes; that is, to improve the chance of retrieving semen through percutaneous epididymal sperm extraction or testicular sperm extraction and concomitantly the probability of conception. Users can click the boxes of the checklist, which are also synchronized with those of their partner.

In addition to the 5 tabs, myFertiCare provides a main menu with general information (eg, contact details and app settings) and a link to the user's personal health record. In the personal health record, users can see their own test results and read the correspondence between their primary care provider and their family physician.

Finally, for couples who are not yet being treated at RUMC and thus do not have login details, a preview version of myFertiCare is available. In this version, they can view the general treatment trajectory and consult the checklist with lifestyle advice. Thus, they can prepare themselves for their intake appointment.

Phase 3: Qualitative Evaluation of myFertiCare for Usability With the Think-Aloud Method

The think-aloud sessions yielded both positive and negative feedback. Given the aim of the study, we focused on opportunities to improve the app. As described earlier, every participant completed 16 tasks (11 generic and 5 personalized). This resulted in 21 different tasks. After each task, the participants answered 3 task-linked questions ("I find this task easy," "I find this information useful," "I find this information is in a logical spot") on a 5-point Likert scale (1="totally disagree," 5="totally agree"). Figure 1 shows a summary of the results. In general, participants considered the tasks (ie, the functionalities that myFertiCare provides) easy and useful. They also considered that the information was provided logically.

Although the scores for all tasks were high, the participants named some discomforts and suggestions for improving the app design. They commented that logging in with their DigiD was too cumbersome, since it consists of a username, password, and verification via text message. It was also noted that moving along the visualized treatment trajectory was difficult. The participants attempted to slide through the treatment trajectory, which was not possible. Instead, they had to click on stops to move to this specific stop. Furthermore, they noticed that the app did not open with the most recent appointment, which was the mode they preferred. When using the forum, participants regretted that they could not delete an erroneous message they had posted. Finally, participants expressed the need for a home button to lead them to the home screen of the app.

At the end of each think-aloud session, the participants completed a questionnaire about their experience using myFertiCare. Figure 2 shows a summary of the results. The participants allocated high scores to all surveyed items that related to usability, understandability of information, the usefulness of various functionalities, and privacy. The men were consistently slightly more critical than the women. The space for writing free comments revealed no additional information. All participants felt that myFertiCare provides an added value to them. All but 1 participant intended to use myFertiCare in the future. In conclusion, myFertiCare was rated 8 out of 10 (Figure 2).

Guided by the think-aloud sessions, we made various improvements in app design. We made it possible for myFertiCare users to create a 4-digit entry code after the first login with DigiD, so that fast but equally safe access was enabled for future use. Furthermore, opening myFertiCare with the most recent appointment was made possible, while proceeding through the treatment trajectory. We added an option to remove a message from the forum after it has been posted as well. A home button was incorporated, which leads users to the app's home screen.

Discussion

Principal Findings

We designed, developed, and qualitatively evaluated an eHealth app for fertility care in accordance with a methodological framework, based on couples' information needs and input provided by health care providers. The basis of the app is the visualized treatment trajectory. The app provides both personalized and interactive functionalities, including customized information and communication options. On thorough qualitative evaluation, myFertiCare received high usability ratings. The participants felt that myFertiCare provides an added value during their treatment. The app was rated with a median score of 8 out of 10. The most important improvements after the think-aloud sessions were related to faster login and easier navigation through the app.

A large part of research in fertility care is aimed at the female partner. We chose to include both partners when developing the app, since it is recognized that men should have a well-defined role as an equal partner during fertility treatment, particularly in cases of male infertility [16]. A previous study by Sylvest et al [11] reported that men registered a marked time lag between diagnosis and treatment initiation. They felt “they were in a maze without a map” and expressed the need for detailed information about the treatment plan, including a timetable, so that they could control and manage their lives [11]. With a visualized treatment trajectory as the basis of myFertiCare, we aimed to meet this need and guide couples through their entire treatment trajectory. In our opinion, patient satisfaction with information provided by the clinic is an important indicator of the quality of fertility care, although in fertility care, the focus is often on live birth rates. Alper et al [17] further endorsed this idea.

There is literature available on eHealth initiatives in fertility care [13], primarily on online support groups. In general, there is a lack of initiatives that provide interactive and dynamic elements, and there is a lack of methodological standards for these complex interventions [13]. There has been 1 web-based initiative that provides both information and peer support, which showed high patient appreciation [18]. Furthermore, a web-based community has been reported in which couples could communicate and share information with professionals and peers [19].

Compared to previous initiatives, a strength of myFertiCare is that it provides a large variety of personalized and interactive functionalities centered around the visualized treatment trajectory of the couple. Another methodological strength of our study is its 3-phase systematic approach: first, a functional design for the app was developed through a qualitative assessment of the informational needs of patients; second, myFertiCare was actually developed; and third, myFertiCare was qualitatively evaluated for usability through the think-aloud method. All 3 phases were carried out in collaboration with patients and medical and technical professionals, which is important for successful eHealth development and implementation [20]. Our qualitative evaluation of myFertiCare for usability is crucial since usability evaluations are critical to the success of adopting an interactive eHealth app [20,21]. The think-aloud method is preferred for uncovering usability problems, and it provides understanding of how users interact with myFertiCare [22]. Furthermore, the think-aloud method is especially suitable since we used both a concurrent method (ie, while performing the task) and a retrospective method (ie, immediately after performing the task) to report participants' thinking, a method that has been suggested for producing optimal data quality [23].

Limitations

Our study also has some limitations. It could be argued that the study population for the think-aloud sessions was relatively small. However, studies have shown that for a think-aloud test, 5 participants are enough for success in identifying usability problems in a user interface [24]. Since we included 9 participants and achieved data saturation, we are confident that we have identified all the possible usability problems. Furthermore, it is known to be a challenge for a researcher to remain consistent when it is necessary to intervene in a think-aloud session; for example, when a participant is unable to complete a task, clarification on a participant's comment is required, or a participant sidesteps the functionality of interest

[25]. In these situations, it is important to explain to a participant that it is the aim of the study to identify problems and to invite them to approach the problem otherwise. It has been reported that researchers often unintentionally intervene in theoretically inconsistent ways [25]. We made a conscious effort to achieve a reliable data set by being aware of these limitations and through triangulation of research methods (namely the think-aloud, task-linked questions, and researcher's observations) and the recording, transcribing, and coding of the interviews.

Practical Implications

This study provides a framework for patient-centered design, development, and evaluation of an eHealth app. Our systematic approach, in which patients and professionals participated in every phase of the process, is particularly suitable in the current era where patient-centeredness is highly valued. Furthermore, we obtained insight into the various functionalities that patients appreciate in a web-based app. The framework we developed for myFertiCare supports professionals in fertility care for guiding patients through their treatment trajectory and delivering patient-centered care. In the near future, myFertiCare will also be evaluated quantitatively. Expansion of eHealth tools to cover the whole fertility care journey and expansion to other medical disciplines is considered of high value. Development of eHealth tools from a patient's viewpoint is an opportunity to empower patients in managing their own treatment trajectories in the current era of patient-centered care.

Conclusions

We designed, developed, and qualitatively evaluated a multifaceted web-based app, myFertiCare, through a systematic approach in which patients and medical and technical professionals participated in every phase. This app aims to guide couples undergoing ICSI with surgically retrieved sperm through their treatment trajectory. myFertiCare provides personalized and interactive functionalities, facilitating the provision of patient-centered care and empowering patients to manage their own treatment trajectory. The app had a high usability rating and was highly appreciated by both male and female partners.

Acknowledgments

The authors wish to thank all study participants and the medical and technical experts, who contributed to the development and evaluation of myFertiCare. This study did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Authors' Contributions

All authors contributed to the design of the study. EMS performed data collection and analysis, which was supervised by WLDMN and KF. EMS drafted the manuscript. All authors critically revised the manuscript and approved the final version for submission for publication.

Conflicts of Interest

None declared.

References

1. Corrigan J, Donaldson M, Kohn L, Maguire S, Pike K. Crossing the quality chasm: A new health system for the 21st Century. Washington, DC: National Academies Press; 2001.
2. Aarts JWM, Huppelschoten AG, van Empel IWH, Boivin J, Verhaak CM, Kremer JAM, et al. How patient-centred care relates to patients' quality of life and distress: a study in 427 women experiencing infertility. Hum Reprod 2012 Feb;27(2):488-495. [doi: 10.1093/humrep/der386] [Medline: 22108249]
3. van Empel IWH, Nelen W, Tepe E, van Laarhoven EAP, Verhaak C, Kremer J. Weaknesses, strengths and needs in fertility care according to patients. Hum Reprod 2010 Jan;25(1):142-149. [doi: 10.1093/humrep/dep362] [Medline: 19861329]
4. Mourad S, Hermens RPMG, Cox-Witbraad T, Grol RPTM, Nelen WLDM, Kremer JAM. Information provision in fertility care: a call for improvement. Hum Reprod 2009 Jun;24(6):1420-1426. [doi: 10.1093/humrep/dep029] [Medline: 19224886]
5. Dancet EAF, Nelen WLDM, Sermeus W, De Leeuw L, Kremer JAM, D'Hooghe TM. The patients' perspective on fertility care: a systematic review. Hum Reprod Update 2010;16(5):467-487. [doi: 10.1093/humupd/dmq004] [Medline: 20223789]
6. van Empel IWH, Aarts J, Cohlen B, Huppelschoten D, Laven J, Nelen W, et al. Measuring patient-centredness, the neglected outcome in fertility care: a random multicentre validation study. Hum Reprod 2010 Oct;25(10):2516-2526. [doi: 10.1093/humrep/deq219] [Medline: 20719811]
7. Dancet E, Van Empel IWH, Rober P, Nelen W, Kremer J, D'Hooghe TM. Patient-centred infertility care: a qualitative study to listen to the patient's voice. Hum Reprod 2011 Apr;26(4):827-833. [doi: 10.1093/humrep/der022] [Medline: 21317152]
8. Dancet EAF, D'Hooghe TM, Sermeus W, van Empel I, Strohmer H, Wyns C, et al. Patients from across Europe have similar views on patient-centred care: an international multilingual qualitative study in infertility care. Hum Reprod 2012 Jun;27(6):1702-1711. [doi: 10.1093/humrep/des061] [Medline: 22427309]
9. Weissman A, Gotlieb L, Ward S, Greenblatt E, Casper RF. Use of the internet by infertile couples. Fertil Steril 2000 Jun;73(6):1179-1182. [doi: 10.1016/s0015-0282(00)00515-x] [Medline: 10856479]

10. Haagen E, Tuil W, Hendriks J, de Bruijn RPJ, Braat DDM, Kremer JAM. Current Internet use and preferences of IVF and ICSI patients. Hum Reprod 2003 Oct;18(10):2073-2078. [doi: 10.1093/humrep/deg423] [Medline: 14507823]
11. Sylvest R, Fürbringer JK, Schmidt L, Pinborg A. Infertile men's needs and assessment of fertility care. Ups J Med Sci 2016 Nov;121(4):276-282 [FREE Full text] [doi: 10.1080/03009734.2016.1204393] [Medline: 27501219]
12. Ball MJ, Lillis J. E-health: transforming the physician/patient relationship. Int J Med Inform 2001 Apr;61(1):1-10. [doi: 10.1016/s1386-5056(00)00130-1] [Medline: 11248599]
13. Aarts J, van den Haak P, Nelen W, Tuil W, Faber M, Kremer JAM. Patient-focused internet interventions in reproductive medicine: a scoping review. Hum Reprod Update 2012;18(2):211-227. [doi: 10.1093/humupd/dmr045] [Medline: 22108381]
14. Sparidaens EM, Braat DD, van den Berg M, D'Hauwers KWM, Fleischer K, Nelen WL. Informational needs of couples undergoing intracytoplasmic sperm injection with surgical sperm retrieval: A qualitative study. Eur J Obstet Gynecol Reprod Biol 2020 Dec;255:177-182. [doi: 10.1016/j.ejogrb.2020.10.040] [Medline: 33166938]
15. Someren VM, Barnard Y, Sanberg J. The Think Aloud Method: A Practical Guide to Modelling Cognitive Processes (volume 12). Cambridge, MA: Academic Press; 1994.
16. Mikkelsen AT, Madsen SA, Humaidan P. Psychological aspects of male fertility treatment. J Adv Nurs 2013 Sep;69(9):1977-1986. [doi: 10.1111/jan.12058] [Medline: 23215849]
17. Alper M, Brinsden P, Fischer R, Wikland M. Is your IVF programme good? Hum Reprod 2002 Jan;17(1):8-10. [doi: 10.1093/humrep/17.1.8] [Medline: 11756353]
18. Tuil W, ten Hoopen AJ, Braat DDM, de Vries Robbé PF, Kremer JAM. Patient-centred care: using online personal medical records in IVF practice. Hum Reprod 2006 Nov;21(11):2955-2959. [doi: 10.1093/humrep/del214] [Medline: 16982658]
19. Aarts JW, Faber MJ, den Boogert AG, Cohlen BJ, van der Linden PJ, Kremer JA, et al. Barriers and facilitators for the implementation of an online clinical health community in addition to usual fertility care: a cross-sectional study. J Med Internet Res 2013 Aug 30;15(8):e163 [FREE Full text] [doi: 10.2196/jmir.2098] [Medline: 23996964]
20. Yen P, Bakken S. Review of health information technology usability study methodologies. J Am Med Inform Assoc 2012;19(3):413-422 [FREE Full text] [doi: 10.1136/amiajnl-2010-000020] [Medline: 21828224]
21. Venkatesh V, Morris MG, Davis GB, Davis FD. User Acceptance of Information Technology: Toward a Unified View. MIS Quarterly 2003;27(3):425. [doi: 10.2307/30036540]
22. Rose AF, Schnipper JL, Park ER, Poon EG, Li Q, Middleton B. Using qualitative studies to improve the usability of an EMR. J Biomed Inform 2005 Feb;38(1):51-60 [FREE Full text] [doi: 10.1016/j.jbi.2004.11.006] [Medline: 15694885]
23. Sudman S, Brandburn N, Schwarz N. Thinking About Answers: The Application of Cognitive Processes to Survey Methodology. Hoboken, NJ: Wiley; 1996.
24. Nielsen J. Estimating the number of subjects needed for a thinking aloud test. Int J Hum Comput Stud 1994 Sep;41(3):385-397. [doi: 10.1006/ijhc.1994.1065]
25. Boren T, Ramey J. Thinking aloud: reconciling theory and practice. IEEE Trans Profess Commun 2000;43(3):261-278. [doi: 10.1109/47.867942]

Abbreviations

DigiD: digital identity
ICSI: intracytoplasmic sperm injection
RUMC: Radboud University Medical Center

Edited by R Kukafka; submitted 30.10.20; peer-reviewed by Y Motoki, C Reis; comments to author 20.12.20; revised version received 07.04.21; accepted 25.04.21; published 03.08.21.

Please cite as:
Sparidaens EM, Hermens RPM, Braat DDM, Nelen WLDM, Fleischer K
Web-Based Guidance Through Assisted Reproductive Technology (myFertiCare): Patient-Centered App Development and Qualitative Evaluation
J Med Internet Res 2021;23(8):e25389
URL: https://www.jmir.org/2021/8/e25389
doi:10.2196/25389
PMID:34342591

©Ellen Marie Sparidaens, Rosella P M Hermens, Didi D M Braat, Willianne L D M Nelen, Kathrin Fleischer. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 03.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet

Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Evaluation of a Digital Health Initiative in Illicit Substance Use: Cross-sectional Survey Study

Steph Kershaw[1], BLabMed (Hons), PhD; Louise Birrell[1], BSocSci, BPsych (Hons), PhD; Hannah Deen[1], BPsych (Hons); Nicola C Newton[1], BPsych (Hons), PhD; Lexine A Stapinski[1], BPsych (Hons), MClinPsych, PhD; Katrina E Champion[1], BAPsych (Hons), BHealth, PhD; Frances Kay-Lambkin[2], BScPsych (Hons), PhD; Maree Teesson[1], BScPsych (Hons), PhD; Cath Chapman[1], BA (Hons), PhD

[1]The Matilda Centre for Research in Mental Health and Substance Use, University of Sydney, Sydney, Australia

[2]Priority Research Centre for Brain and Mental Health, The University of Newcastle, Newcastle, Australia

Corresponding Author:
Steph Kershaw, BLabMed (Hons), PhD
The Matilda Centre for Research in Mental Health and Substance Use
University of Sydney
Level 6, Jane Foss Russell building
Camperdown
Sydney, 2006
Australia
Phone: 61 2 8627 9018
Email: steph.kershaw@sydney.edu.au

Abstract

Background: The *Cracks in the Ice* (CITI) community toolkit was developed to provide evidence-based, up-to-date information and resources about crystal methamphetamine to Australians. Given the high rates of internet use in the community and the potential for misinformation, CITI has the potential to play an important role in improving knowledge and challenging misconceptions surrounding crystal methamphetamine.

Objective: This study aims to determine (1) whether the CITI toolkit is achieving its aim of disseminating evidence-based information and resources to people who use crystal methamphetamine, their family and friends, health professionals, and the general community and (2) examine the association between the use of CITI and the knowledge and attitudes about crystal methamphetamine.

Methods: A cross-sectional web-based survey, open to Australian residents (aged ≥18 years), was conducted from November 2018 to March 2019. People who had previously visited the website (referred to as "website visitors" in this study) and those who had not ("naïve") were recruited. At baseline, knowledge, attitudes, and demographics were assessed. CITI website visitors then completed a series of site evaluation questions, including the System Usability Scale (SUS), and naïve participants were asked to undertake a guided site tour of a replicated version of the site before completing the evaluation questions and repeating knowledge and attitude scales.

Results: Of a total 2108 participants, 564 (26.7%) reported lifetime use of crystal methamphetamine, 434 (20.6%) were family/friends, 288 (13.7%) were health professionals, and 822 (38.9%) were community members. The average SUS score was 73.49 (SD 13.30), indicating good site usability. Health professionals reported significantly higher SUS scores than community members (P=.02) and people who used crystal methamphetamine (P<.001). Website visitors had significantly higher baseline knowledge than naïve participants (P<.001). Among naïve participants, knowledge scores increased following exposure to the website (mean 15.2, SE 0.05) compared to baseline (mean 14.4, SE 0.05; P<.001). The largest shifts in knowledge were observed for items related to prevalence, legal issues, and the effects of the drug. Stigmatizing attitude scores among the naïve group were significantly lower following exposure to CITI (mean 41.97, SE 0.21) compared to baseline (mean 44.3, SE 0.21; P<.001).

Conclusions: This study provides an innovative evaluation of a national eHealth resource. CITI is achieving its aim of disseminating evidence-based, nonstigmatizing, and useful information and resources about crystal methamphetamine to key end user groups and has received good usability scores across its target groups. Interaction with CITI led to immediate improvements in knowledge about crystal methamphetamine and a decrease in stigmatizing attitudes. CITI demonstrates the important role of digital information and support platforms for translating evidence into practice and improving knowledge and reducing stigma.

(J Med Internet Res 2021;23(8):e29026) doi:10.2196/29026

KEYWORDS

methamphetamine; eHealth; substance-related disorders; internet; preventive psychiatry; health education; mobile phone

Introduction

Background

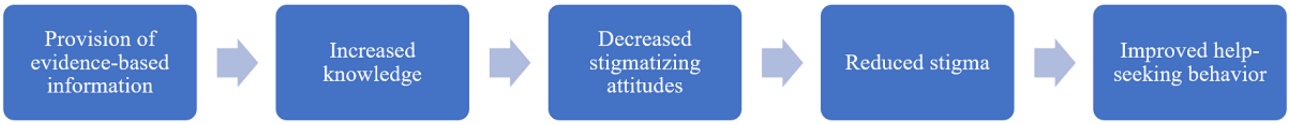
Internationally, there continues to be a widespread concern about the use of methamphetamine, particularly with respect to the most potent form, crystal methamphetamine. Crystal methamphetamine produces stronger and longer-lasting effects than other forms of methamphetamine, thereby increasing the risk of serious long-term health issues, including dependence [1,2]. The use of crystal methamphetamine has been associated with considerable harms not only in individuals who use the drug but also in their families and communities [3]. Crystal methamphetamine has attracted a high level of negative attention compared with other drugs, particularly in Australia. In the 2019 National Drug Strategy Household Survey, the Australian community ranked meth/amphetamines as the drug of most serious concern and the drug most likely to be associated with a *drug problem* [4]. Its public ranking overtook alcohol in 2016 for the first time in the history of the survey [5]. Concerningly, crystal methamphetamine use has been frequently labeled as a national *crisis* or *epidemic*, with people who use crystal methamphetamine being negatively stereotyped in the media and being framed as criminal, deviant, or dangerous [6]. This is despite the prevalence of methamphetamine use being low, with only 1% of Australians reporting use in the previous 12 months [4].

In response to growing concern and negative attention around the drug, the Australian Government's Department of Health funded the development of the *Cracks in the Ice* (CITI) community toolkit as an easily accessible, evidence-based, and up-to-date information resource [7]. The website was launched in 2017 and includes information about the prevalence, effects of the drug, treatment options, and support programs as well as tailored resources for families or friends, health professionals, and the community. The codevelopment process of CITI was broad-reaching and iterative, with the involvement of more than 450 community members across Australia (all states and territories), including people with lived experience, their families and friends, health professionals, and researchers [8]. CITI currently has 107 resources including factsheets, guidelines, animations or videos, and training and support programs and has reached more than 625,000 unique end users from across the globe (70% from Australia) with more than 1.2 million page views (Google Analytics as of March 22, 2021) and 2 national awards (Rotary Health Australia, 2017; The Mental Health Services Learning Network, 2020). A companion smartphone app, which allows users to access content offline, was launched in January 2018 to extend the reach and engagement of the toolkit, particularly for people in areas where internet access may be unreliable (ie, remote or regional areas) [9]. This is essential in a geographically diverse country such as Australia, where several data sources have indicated that the use of methamphetamine is higher, access to face-to-face services is lower, and unreliable internet connections are more common in regional and remote areas than in urban areas [10-12].

The internet has become a leading source of health information for the public, with eHealth platforms such as CITI offering a high level of accessibility and potential reach [13]. eHealth programs and websites have demonstrated an increase in empowerment and knowledge among health seekers on the internet [14] along with improved quality of care through the translation of evidence-based research into practice [15,16]. In Australia, it is estimated that 86.5% of the population are internet users, and 80% of Australians reported using the internet to search for health information in 2019 [17]. Indeed, during the development of CITI, the most commonly endorsed reasons given by participants for visiting a website about crystal methamphetamine were to "seek information for myself" (41.0%) and "to find out how to get help for a friend or a family member" (30.2%) [8]. However, there are risks associated with using the internet as a source of information, with many websites containing low-quality, irrelevant, inaccurate, or inappropriate health information [18,19]. Thus, a rigorous methodology is needed for both the development and evaluation of websites and eHealth resources. However, a recent systematic review found that although there are many toolkits for digital health, evaluations of these toolkits are uncommon [20]. This study aims to fill this gap. This evaluation is also one of the first to examine the relationship between knowledge and its impact on attitudes, an element not often looked at in website evaluations.

Given the large proportion of Australians who use the internet, the potential for misinformation, and the high level of interest in crystal methamphetamine use, initiatives such as CITI play an important role in not only disseminating accurate evidence-based information but also reducing the stigma and discrimination surrounding crystal methamphetamine use. Recent research by our team indicates that stigma toward people who use crystal methamphetamine is common in Australia, with 1 in 3 people who use crystal methamphetamine reporting that they have been discriminated against because of their drug use [21]. Stigma is a source of immense psychological distress and has been associated with feelings of self-blame, low self-worth, shame, and higher rates of drug dependence and lower rates of treatment seeking for substance use problems and other health conditions such as mental illness, trauma, and infections [22-24]. Importantly, our previous study also demonstrated that higher levels of accurate knowledge about crystal methamphetamine were associated with less stigmatizing attitudes [21]. Thus, by improving knowledge about crystal methamphetamine, CITI has the potential to influence public attitudes toward those affected by the drug and support people to seek help when they need it (Figure 1).

Figure 1. Proposed conceptual model of how providing evidence-based information can lead to increased knowledge and decreased stigma and ultimately improve help-seeking behaviors among people who use crystal methamphetamine.


Provision of evidence-based information
Increased knowledge
Decreased stigmatizing attitudes
Reduced stigma
Improved help-seeking behavior

Objectives

This study presents a large community evaluation of an innovative centralized eHealth resource for evidence-based information and resources on crystal methamphetamine. The aims of this study were to (1) determine whether the CITI toolkit is achieving its aim of disseminating evidence-based, nonstigmatizing, and useful information and resources about crystal methamphetamine to key end user groups (people who use crystal methamphetamine, affected family members and friends, health professionals, and the general Australian community) and (2) examine the association between the use of CITI and improvements in knowledge and attitudes about crystal methamphetamine and harms.

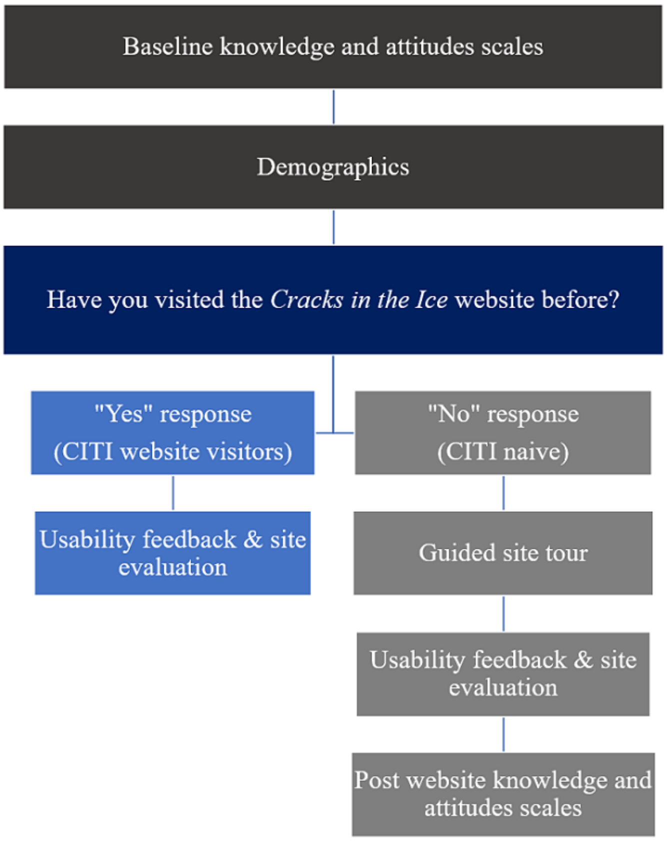
Methods

Design and Procedure

Between November 2018 and March 2019, participants were recruited via convenience sampling methods, including advertising on the CITI website, Facebook page, Twitter page, e-newsletter, and through paid public advertising on Facebook. The eligible participants provided informed consent before completing a web-based cross-sectional survey that was open to all Australian residents aged ≥18 years. At the end of the survey, all participants were given the opportunity to provide their email address to enter a draw to win an Apple iPad. Ethics approval was obtained from the University of New South Wales (HC# HC180735) and the University of Sydney (Project 2018/844) human research ethics committees.

To address both aims of the study, people who had previously visited the website (CITI website visitors) and people who had not previously visited the website (CITI naïve) were recruited (Figure 2). At baseline, all participants completed a series of knowledge and attitude scales, along with standard demographic questions. Participants were then divided into 2 groups depending on their answer to the question “Have you visited the CITI website before?” (yes, no, or unsure). CITI website visitors completed a series of site evaluation questions, whereas CITI-naïve participants were asked to undertake a guided site tour of a replicated version of the live CITI site before completing the site evaluation questions and repeating the knowledge and attitude scales. This methodology permitted the assessment of any change in these constructs associated with viewing the website. As a part of the guided site tour, CITI-naïve participants were encouraged to view the homepage and key webpages featuring content on the prevalence and physical and mental health effects of crystal methamphetamine use along with treatment and support options. As CITI is a live responsive website, the entire site was replicated to allow measurement of usage behavior by study participants. Participants who selected *unsure* were assigned to the CITI-naïve group; however, they were excluded from group comparison analyses to reduce any possible confounding of results.

Figure 2. Flow chart of the study design. CITI: *Cracks in the Ice*.


Baseline knowledge and attitudes scales
Demographics
Have you visited the *Cracks in the Ice* website before?
"Yes" response (CITI website visitors)
"No" response (CITI naive)
Usability feedback & site evaluation
Guided site tour
Usability feedback & site evaluation
Post website knowledge and attitudes scales

Measures

Information collected included age (in years), gender, residential postcode, and region (metropolitan, regional, or rural or remote). Participants were asked to identify whether they had previously used crystal methamphetamine, had a family member or friend who used (or who they thought might be using) crystal methamphetamine, or were a health professional. All participants were asked to indicate whether they had previously viewed the CITI website.

The System Usability Scale (SUS) is a reliable scale for measuring usability, including 10 items about several facets such as the complexity of the website and ease of use. The 10 items were scored on a 5-point Likert scale ranging from 0 (strongly disagree) to 5 (strongly agree), with scores calculated as validated in the literature [25]. Scale scores ranged from 0 to 100, with higher scores indicating enhanced usability. The average SUS score based on more than 500 studies was 68, with scores of 80 or above indicating a strong performance [26].

Participants also responded to an additional 4 items about the website's purpose, goals, and evidence base adapted from previous research by our team [8] with items being (1) "The support options listed on *Cracks in the Ice* are useful," (2) "The primary goal/purpose of *Cracks in the Ice* is clear," (3) "The information and resources on *Cracks in the Ice* have been informed by evidence," and (4) "The terminology on *Cracks in the Ice* is nonstigmatizing." These statements were scored on a 5-point Likert scale, ranging from 0 (strongly disagree) to 5 (strongly agree). In addition, participants were asked on a 7-point Likert scale, ranging from 0 (strongly disagree) to 7 (strongly agree), "I am confident in helping someone who has a problem with ice" and on a 5-point Likert scale, "If I knew someone who was using ice, I would help them." Participants were given the opportunity throughout the questionnaire to provide open-ended feedback about the website. The term "ice" was used instead of "crystal methamphetamine" as it is a more commonly known term for crystal methamphetamine among the Australian community.

To assess knowledge of crystal methamphetamine, participants completed a series of 18 *true or false* questions (each with three response options: 1=true, 2=false, and 3=unsure). Questions included knowledge of its effects on the brain and body (eg, "Ice will cause psychosis in all people"), rates of use (eg, "Methamphetamines [including ice] are the most popular illicit drugs in Australia"), and legal status in Australia. The total knowledge score was calculated by summing the number of correctly answered items (min=0 and max=18). *Unsure* responses were coded as incorrect for the purposes of analysis. This measure was developed by the authors as part of the development of CITI to assess key areas of knowledge and potential misconceptions about crystal methamphetamine among a lay audience [8] and has been used in other analyses by our group [21].

To assess attitudes toward people who use methamphetamine, participants responded to a 14-item, 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Ten items were adapted from the Brener and Von Hippel scale to measure attitudes toward people who inject drugs. The reliability and validity of the original scale have been previously demonstrated [27]. Mentions of "people who inject drugs" in the original scale were replaced with "people who use ice." Four additional items were adapted from the Depression Stigma Scale (DSS) with mention of mental health issues being replaced with "people who use ice" to assess perceptions of personal blame or control ("A person using ice could recover if they just stopped using" or "Using ice is a sign of personal weakness"), danger ("People who use ice are dangerous"), and shame ("If I had a problem with ice, I would not tell anyone") [28]. Although the DSS was originally designed to measure stigma attached to depression, it has also been used to assess stigma attached to other disorders [29,30]. The reliability and validity of the DSS have been demonstrated in prior studies [28,30]. Positive statements were reverse coded before the total score was calculated (min=14 and max=70), with higher scores indicating more negative attitudes. The adapted scale demonstrated good internal consistency in the current sample (Cronbach α=.86).

Data Analysis

Data were analyzed using the IBM SPSS Statistics version 24. Analysis of variance was conducted to assess differences in system usability scores of end user groups and the impact of CITI on knowledge and attitudes about crystal methamphetamine. Chi-square tests were used to compare ratings on Likert scale statements among the 4 end user groups, with "strongly disagree" and "disagree" ratings combined for the purpose of analyses, as the count was too low among the health professional end user group. Two-tailed *t* tests were also conducted to assess whether knowledge was different between the CITI website visitors and CITI-naïve groups and within the CITI-naïve group before and after visiting the CITI website. Content analysis of open-ended feedback using general inductive analysis [31] was also performed to identify core themes.

Results

Participants

A total of 2125 participants completed the survey. Among them, 17 participants were excluded from the analysis owing to invalid responses (eg, rapid responding or inconsistent responses), leaving a total of 2108 participants in the final sample. The mean age was 36 (SD 13) years, and 58.6% (1235/2108) of the participants identified as female. More than half of the participants were from metropolitan areas (1190/2108, 56.5%), a third (691/2108, 32.8%) were from regional locations, and 10.8% (227/2108) were from rural or remote locations. The sample's age, gender, and state or territory of residence aligned closely with national population statistics from the 2018 Australian Bureau of Statistics [32]. A total of 26.7% (564/2108) of participants reported lifetime use of crystal methamphetamine (any use of the drug either current or in the past). Table 1 summarizes the descriptive statistics of gender, age, and region for each of the target end user groups (people who use crystal methamphetamine, affected family and friends, health professionals, and the general Australian community). The respondents in the "people who use crystal methamphetamine" group included both people who were currently using and those who reported using it in the past.

Table 1. Demographics by end user groups (N=2108).

Demographics	End user group			
	People who use crystal methamphetamine (n=564)	Family and friends (n=434)	Health professionals (n=288)	Community members (n=822)
Gender, n (%)				
Female	278 (49.3)	306 (70.5)	212 (73.6)	439 (53.4)
Male	270 (47.9)	113 (26.0)	75 (26.0)	368 (44.8)
Nonbinary or gender fluid	5 (0.9)	11 (2.5)	0 (0)	13 (1.6)
Different identity or prefer not to say	11 (1.9)	4 (1.0)	1 (0.3)	2 (0.2)
Age (years), mean (SD)	35.76 (9.42)	39.45 (14.55)	40.42 (12.72)	33.57 (13.64)
Region, n (%)				
Metro	343 (60.8)	215 (49.5)	146 (50.7)	486 (59.1)
Regional	176 (31.2)	161 (37.1)	92 (31.9)	262 (31.9)
Rural or remote	45 (8.0)	58 (13.4)	50 (17.4)	74 (9.0)

Usability and Site Evaluation

The average SUS score for the CITI website was 73.49 (SD 13.30), indicating good usability. Table 2 provides the average SUS score for each end user group. SUS scores were significantly higher in the health professionals group than in the community ($F_{3,2062}$=6.8; P=.02) and people who use crystal methamphetamine ($F_{3,2062}$=6.8; P<.001) end user groups. SUS scores were also significantly higher in the family and friends group than in the people who used crystal methamphetamine group ($F_{3,2062}$=6.8; P=.04).

Table 2. System Usability Scale scores for each end user group (N=2066)[a].

End user group	Participants, n (%)	SUS[b], mean (SD)
Community members	809 (39.1)	73.53 (13.43)
People who use crystal methamphetamine	553 (26.7)	71.73 (13.31)
Family and friends	419 (20.2)	74.12 (13.92)[c]
Health professionals	285 (13.7)	75.91 (11.41)[d]

[a] A total of 42 participants were excluded from the analyses because they selected unsure when asked if they had visited the *Cracks in the Ice* website.

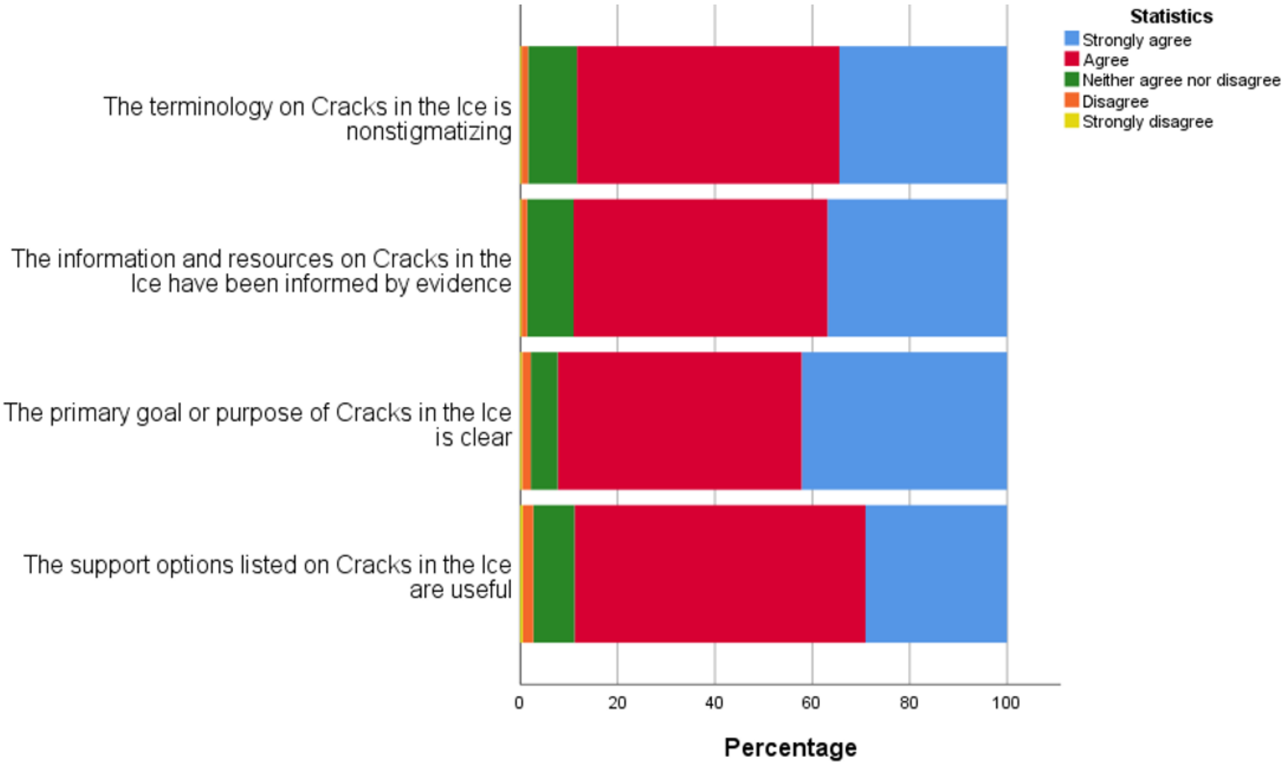
[b] SUS: System Usability Scale.

[c] P=.02.

[d] P<.01.

Participants had an overall positive response to the website's purpose, goals, and evidence base (Figure 3), with most participants providing a rating of *agree* or *strongly agree* for statements such as "The information and resources has been informed by evidence" (1839/2066, 87.2%), "The terminology is nonstigmatizing (1825/2066, 86.6%), "The primary goal/purpose is clear" (1907/2066, 90.5%), and "The support options listed are useful" (1836/2066, 87.1%).

Figure 3. Participant feedback of the *Cracks in the Ice* website's purpose, goals and, evidence base on a 5-point Likert scale.



Ratings for each statement were compared, with significant associations observed between the end user group and ratings for each of the statements (χ^2_9 range 43.79-67.05; P<.001). Agreement was generally higher among the community, health professional, and family and friends groups compared with people who use crystal methamphetamine; however, the differences were small (between 5% and 12%). Further descriptive information for each statement by the end user group is detailed in Multimedia Appendix 1.

Open-ended comments left by survey respondents at the end of the survey were mostly positive, stating that the website was evidence-based, informative, easy to use, and presented information in a nonjudgmental or nonstigmatizing way. Other comments related to improving the user experience or navigation, creating more resources such as lived experience stories, and specific resources to support regional or rural areas along with Aboriginal and Torres Strait Islander peoples. Several examples of the comments are provided in Multimedia Appendix 2.

Knowledge and Attitudes About Crystal Methamphetamine and Harms

Overview

There were 277 participants who had previously visited the website (CITI website visitors) and 1789 participants who had not previously visited the website (CITI-naïve group). An additional 42 participants were excluded from further analyses as they selected the response *unsure* when asked if they had visited the website.

The CITI-naïve group spent an average of 2.16 minutes on the replicated website, visiting 2.17 pages, spending an average of 82.63 (SD 189.5) seconds on relevant content pages (Google Analytics November 19, 2018, to March 31, 2019). This is in line with the average session duration of between 2 and 3 minutes for large websites, such as many health care sites [33].

Previous experience with the website resulted in significantly higher SUS scores in the CITI website group (t_1=8.36; P=.004) compared with the CITI-naïve group.

On average, CITI website visitors had higher total baseline knowledge scores (mean 15.2, SE 0.1) than those who had not previously visited the website (mean 14.4, SE 0.05) (Table 3). This difference was significant (t_{2064}=5.54; P<.001), with a small to medium–sized effect (Cohen d=0.44).

Table 3. Comparison of knowledge between *Cracks in the Ice* (CITI) visitors and the CITI-naïve group at baseline, indicating the numbers and percentages of participants who gave correct answers to the knowledge statement (N=2108)[a].

Knowledge statement	Total sample (n=2108), n (%)	CITI[b] visitors (n=277), n (%)	CITI-naïve (n=1789), n (%)
It is illegal to drive while under the influence of ice	2069 (98.1)	271 (97.8)	1756 (98.15)
Ice use can lead to serious long-term physical effects	2056 (97.5)	270 (97.5)	1745 (97.54)
Using ice can affect your sleep	2019 (95.8)	273 (98.6)	1705 (95.30)
Ice use can make you feel paranoid	2017 (95.7)	272 (98.2)	1705 (95.30)
Ice use can make you feel agitated or distressed	2013 (95.5)	272 (98.2)	1701 (95.08)
You always know what you're taking when you use ice	1962 (93.1)	259 (93.5)	1663 (92.95)
Taking ice with other drugs can reduce the risks of harm	1910 (90.6)	253 (91.3)	1620 (90.55)
Low doses of ice will not impair driving skills	1908 (90.5)	258 (93.1)	1611 (90.05)
It is legal to share ice with friends	1818 (86.2)	238 (85.9)	1544 (86.30)
It is impossible to break an ice dependence	1771 (84.0)	249 (89.9)	1490 (83.28)
The effects of ice are short-lived	1731 (82.1)	235 (84.8)	1464 (81.83)
Ice is generally more potent (strong) than speed	1718 (81.5)	239 (86.3)	1445 (80.77)
You can go to jail for using methamphetamines, including ice	1578 (74.9)	215 (77.6)	1334 (74.56)
Most teenagers have used ice	1570 (74.5)	220 (79.4)	1325 (74.06)
Ice will cause psychosis in all people	1408 (66.8)	216 (78.0)	1172 (65.51)
Legal penalties for drug offences relating to methamphetamine in Australia are different in each of the states and territories	1310 (62.1)	196 (70.8)	1087 (60.76)
Methamphetamines (including ice) are the most popular illicit drugs in Australia	899 (42.6)	144 (52.0)	743 (41.53)
Most of the time, police will not be called when an ambulance is attending a drug overdose	808 (38.3)	126 (45.5)	668 (37.33)

[a] Statements are ordered from most to least correctly answered among the total sample.

[b] CITI: *Cracks in the Ice*.

Change in Knowledge and Attitudes After Viewing the Website

The CITI-naïve group was assessed on knowledge both before and after website viewing, with results showing that, on average, knowledge scores significantly increased following exposure to the website (mean 15.2, SE 0.05) compared with baseline (mean 14.4, SE 0.05; t_{1788}=17.893; P<.001), as displayed in Table 4. The largest shifts in knowledge were observed for items related to the prevalence of crystal methamphetamine use, legal issues, and the physical and mental effects of the drug. There was a 35.1% change (n=261) for "Methamphetamines (including ice) are the most popular illicit drugs in Australia," a 24.0% change (n=160) for "Most of the time, police will not be called when an ambulance is attending a drug overdose," 22.0% change (n=239) for "Legal penalties for drug offences relating to methamphetamine in Australia are different in each of the states and territories," 16.8% change (n=197) for "Ice will cause psychosis in all people," and 10.8% change (n=161) for "It is impossible to break an ice dependence" following exposure to the CITI website.

Table 4. Knowledge among the *Cracks in the Ice*–naïve sample at baseline and after website exposure, indicating the number and percentage of people giving correct answers to knowledge statements (N=1789).

Knowledge statement	Baseline, n (%)	Postexposure[a], n (%)
Methamphetamines (including ice) are the most popular illicit drugs in Australia	743 (41.53)	1004 (56.12)
Most teenagers have used ice	1325 (74.06)	1418 (79.26)
Ice is generally more potent (strong) than speed	1445 (80.77)	1561 (87.25)
You always know what you’re taking when you use ice	1663 (92.95)	1702 (95.13)
Using ice can affect your sleep	1705 (95.30)	1759 (98.32)
Taking ice with other drugs can reduce the risks of harm	1620 (90.55)	1660 (92.78)
Ice use can lead to serious long-term physical effects	1745 (97.54)	1757 (98.21)
It is impossible to break an ice dependence	1490 (83.28)	1651 (92.28)
You can go to jail for using methamphetamines, including ice	1334 (74.56)	1342 (75.01)
It is legal to share ice with friends	1544 (86.30)	1473 (82.33)
Most of the time, police will not be called when an ambulance is attending a drug overdose	668 (37.33)	828 (46.28)
It is illegal to drive while under the influence of ice	1756 (98.15)	1732 (96.81)
The effects of ice are short-lived	1464 (81.83)	1542 (86.19)
Low doses of ice will not impair driving skills	1611 (90.05)	1647 (92.06)
Ice will cause psychosis in all people	1172 (65.51)	1369 (76.52)
Ice use can make you feel agitated or distressed	1701 (95.08)	1761 (98.43)
Ice use can make you feel paranoid	1705 (95.30)	1756 (98.15)
Legal penalties for drug offences relating to methamphetamine in Australia are different in each of the states and territories	1087 (60.76)	1326 (74.11)

[a] Average knowledge scores increased following exposure to the website (mean 15.2, SE 0.05) compared to the baseline (mean 14.4, SE 0.05; t_{1788}=17.893; *P*<.001).

On average, the total attitude score among the CITI-naïve group was significantly lower following exposure to the CITI website (mean 41.97, SE 0.21) compared with the baseline (mean 44.3, SE 0.21; t_{1788}=23.91; *P*<.001), with lower scores indicating less negative or stigmatizing attitudes (Table 5). Follow-up analyses indicated that significant before and after changes in knowledge and attitude scores were observed in all 4 end user groups.

Table 5. Endorsement of attitudes toward people who use crystal methamphetamine of the *Cracks in the Ice*–naïve sample at baseline and after website exposure (N=1789)[a].

Attitude statement	Baseline, n (%)	Postexposure[b], n (%)
Negative attitudes		
I won’t associate with people who use ice if I can help it	1233 (68.92)	986 (55.11)
I avoid people who use ice whenever possible	1232 (68.88)	1103 (61.65)
Use of ice is just plain wrong	986 (55.11)	863 (48.23)
People who use ice are dangerous	821 (45.89)	588 (32.86)
A person using ice could recover if they just stopped using	737 (41.19)	621 (34.71)
If I had a problem with ice, I would not tell anyone	597 (33.37)	395 (22.07)
Use of ice is immoral	470 (26.27)	408 (22.80)
Using ice is a sign of personal weakness	251 (14.03)	182 (10.17)
People who use ice should be locked up to protect society	241 (13.47)	173 (9.67)
Positive attitudes		
People should feel sympathetic and understanding of people who use ice	821 (45.89)	924 (51.64)
People who use ice are mistreated in our society	716 (40.02)	837 (46.78)
People who use ice have a perfect right to their lifestyle, if that's the way they want to live	260 (14.53)	237 (13.24)
People who use ice should be accepted completely into our society	251 (14.03)	317 (17.71)
The use of ice is merely a different kind of lifestyle that should not be condemned	185 (10.34)	186 (10.39)

[a] Participant selected *agree* or *strongly agree* on 5-point Likert scale.

[b] Total attitude score was significantly lower following exposure to the *Cracks in the Ice* website (Mean 41.97, SE 0.21) compared with the baseline (mean 44.3, SE 0.21; t1788=23.91; *P*<.001).

In addition, among the CITI-naïve group, exposure to the CITI website resulted in minor improvements (eg, selecting neutral or agree responses) to the two additional statements “If I knew someone who was using ice, I would help them” and “I am confident in helping someone who has a problem with ice” (Table 6).

For the “If I knew someone who was using ice, I would help them” statement, measured on a 7-point Likert scale from strongly disagree to strongly agree, 25.1% (n=448) increased their rating, 16.3% (n=292) decreased their rating, whereas 58.6% (n=1049) made no change in their rating after viewing the website. For the statement “I am confident in helping someone who has a problem with ice,” measured on a 5-point Likert scale from strongly disagree to strongly agree, 44.1% (n=790) reported an increase in rating, 3.6% (n=64) decreased their rating, whereas 52.3% (n=935) made no change in their rating after viewing the website.

Table 6. Rates of agreement with helping statements before and after website exposure among *Cracks in the Ice*–naïve group (N=1789).

Statement	Baseline, n (%)	Postexposure, n (%)
If I knew someone who was using ice, I would help them		
Strongly disagree	33 (1.8)	30 (1.7)
Mostly disagree	47 (2.6)	47 (2.6)
Somewhat disagree	65 (3.6)	33 (1.8)
Neither agree nor disagree	151 (8.4)	140 (7.8)
Somewhat agree	490 (27.4)	479 (26.8)
Mostly agree	561 (31.4)	543 (30.4)
Strongly agree	442 (24.7)	517 (28.9)
I am confident in helping someone who has a problem with ice		
Not at all	543 (30.4)	191 (10.7)
A little bit	448 (25.0)	448 (25.0)
Moderately	451 (25.2)	596 (33.3)
Quite a bit	234 (13.1)	399 (22.3)
Extremely	113 (6.3)	155 (8.7)

Discussion

Principal Findings

This study evaluated CITI, the first centralized Australian eHealth resource to provide information and resources about crystal methamphetamine to support the community [8]. This evaluation is one of the most rigorous evaluations of eHealth resources to date, conducted among a large sample of more than 2000 Australians. This website evaluation was also one of the first studies to examine the effect of an eHealth resource on knowledge and attitudes regarding illicit substance use. Overall, the results indicated that CITI is achieving its aim of disseminating evidence-based, nonstigmatizing information and resources about crystal methamphetamine to key end user groups (people who use the drug, affected family members and friends, health professionals, and the general Australian community). This is evidenced by more than 80% of participants endorsing statements about the clarity of the website's purpose, goals, and evidence base. Endorsement of the website's purpose and goals was generally higher among the community, health professional, and family and friends groups compared with people who use crystal methamphetamine; however, differences were small (between 5% and 12%). It is a challenge for CITI to cater to different end user groups that have differing needs and has been addressed by providing user-specific sections in addition to general information. Therefore, it was important to check whether these key end user groups had been adequately catered for and identify any improvements that might be needed for the tailored resources.

The average SUS score for CITI was 73.49 (SD 13.30), indicating good usability, which is above the average of 68, based on >500 studies [26]. An SUS of 80 or above is an indication of strong performance, and future iterations of CITI will aim to further improve user experience on the site to reach this level. The differences in SUS scores between different stakeholder groups is a key point of interest, with health professionals reporting significantly higher SUS scores than general community members or people who use crystal methamphetamine. This may be because of health professionals having high digital literacy levels [34] and having higher general levels of education. We also observed higher SUS scores among family and friends than among people who used crystal methamphetamine. This may reflect the fact that the support of families and friends was a priority area in site development, with fewer resources available for people who use crystal methamphetamine. Creating and tailoring information and resources for people who use crystal methamphetamine is an area of growth for the website, and several such resources are currently in development. Unsurprisingly, familiarity with the website (measured by visiting the website before the evaluation) was associated with significantly higher SUS scores.

The finding that the use of CITI led to improvements in knowledge and attitudes about crystal methamphetamine and harms in the naïve group, despite the average time spent on the replicated website being 2 minutes, is encouraging. The average knowledge scores significantly increased after viewing the site, although the effects were small. Examination of individual items showed some promising shifts, particularly for items that are often common misconceptions about the drug (eg, prevalence of use, legal issues, and the physical and mental effects of the drug). Thus, providing some support that the shifts could be meaningful in areas where knowledge is lacking or incorrect. These changes in knowledge are of particular salience given the misconceptions often portrayed in the Australian media about the methamphetamine *epidemic* and the framing of people who use the drug as criminal, deviant, or dangerous [6]. Unsurprisingly, those already familiar with the website before the study had higher total baseline knowledge scores. This may reflect the fact that they had previously reviewed the key informational pages or the possibility that those who had sought out the CITI website of their own volition previously had more knowledge about the drug.

The decrease in total attitude score among the CITI-naïve group (indicating less negative or stigmatizing attitudes) following exposure to the website is also promising. We have previously found that increased knowledge about the drug is associated with less stigmatizing attitudes [21]. This also aligns with what has been found in mental health research, where improvements in knowledge have led to increased positive attitudes or less stigmatizing beliefs [35-38]. Through CITI's dissemination of evidence-based information about crystal methamphetamine, it has the potential to change public attitudes and reduce the stigma associated with the drug. This is potentially an important shift, given that stigma is consistently cited as one of the main barriers to care among people who use methamphetamine [39].

Comparison With Previous Work

eHealth resources are a low-cost way to increase public awareness of important issues and have the potential to reach a large number of people [40]. However, they must respond to rapidly changing population trends and evidence. In areas such as new and emerging drugs, this is particularly important [41]. Therefore, it is important for eHealth resources to be living resources that are adequately funded to respond to emerging research evidence. Consumer health websites need to be not only evidence-based and trustworthy sources of information but also user friendly and meet the health information-seeking needs of their target groups [42,43]. Therefore, website evaluations commonly use a variety of methodologies, including measures of usability (SUS and task performance), self-reports (surveys), qualitative interview-based methods, or Google Analytics to investigate website behaviors (use and interaction) [44-47]. Similar to a recent study, we did not find published recommendations for procedures to conduct a review of a live website in a systematic manner [48]. A recent systematic review found that despite there being many published toolkits, very few were evaluated, and the authors concluded that this is an area where greater attention and rigorous methodology are needed [20]. Although review criteria [49] and a possible framework (Reach, Effectiveness, Adoption, Implementation, Maintenance) from implementation science [50] have been proposed as useful avenues for evaluation, they are more appropriate for web-based intervention–based studies on behavior change. Within this context, we note that this website evaluation is one of the most rigorous evaluations in the field and is a solid starting point, with the potential for future experimental designs. This evaluation also involved the examination of knowledge and its impact on attitudes, an element not often looked at in website evaluations.

It is generally assumed that the longer a visitor spends on a website, the more engaged they are. However, it is difficult to find or estimate a benchmark for the average time spent on health-related websites, as time spent visiting sites depends on the scope and purpose of the actual site. It has been reported that for large websites, such as many health care sites, the average session duration is between 2 and 3 minutes [33]. However, it is difficult to find a benchmark for comparison with public health websites such as CITI because of the large variability across sites and industries. In this study, the CITI-naïve group that undertook the site tour spent an average time of 2.16 minutes, visiting 2.17 pages, in line with the average session duration of health care sites. This is slightly higher than the average time for website users of the live website, which has an average session time of 1.42 minutes with people visiting an average of 1.65 pages (April 3, 2017, to December 31, 2020). It would be useful for future studies to examine the impact of time spent on the site and to determine whether strategies to increase the length of time users engage with the site would be useful to encourage knowledge or attitude change.

The CITI eHealth resource also caters to several different end user populations (eg, health professionals, community members, family and friends, and people who use crystal methamphetamine). As such, similar to other websites, challenges remain in distinguishing between and meeting the needs of different groups [40]. This was evident in this study with people who use crystal methamphetamine having slightly lower endorsement of the website's purpose and goals. Therefore, more in-depth investigation is needed to understand the information needs of people who use crystal methamphetamine. Another avenue for future evaluation is to ensure that the information provided on CITI is suitable to people from diverse backgrounds and varying levels of health literacy, an important consideration that has been noted in the literature [51].

Limitations

This study is not without its limitations, and its findings should be interpreted with these in mind. As CITI is a live website, which can be accessed at any time, this study used a quasi-experimental design. Future evaluations may consider creating beta versions of the site that could be tested in an A/B test with a control group. Another limitation is that although the CITI-naïve group was directed to specific pages during the site tour, there was no minimum time to stay on each page enforced, and participants could also review pages other than those specifically mentioned. Therefore, although changes in knowledge were significant, it is difficult to ascertain the direct cause of this change and whether, for example, it was related to the amount of time participants spent on the site or the specific pages they visited. Similarly, as this study is not longitudinal in its design, it is unclear whether these knowledge shifts are maintained over time or whether changes in knowledge and attitudes lead to behavior change.

This study used convenience sampling methods; therefore, the sample was not nationally representative. However, the sample's age, gender, and state or territory of residence aligned closely with national population statistics from the Australian Bureau of Statistics [32]. Similarly, the use of paid Facebook advertising may have attracted individuals with an interest in the drug crystal methamphetamine; thus, the knowledge and attitude findings may not be representative of the general population.The use of paid advertising also resulted in the sample of CITI-naïve participants being 6 times larger than the group already familiar with CITI; however, the sample sizes achieved in each group were large for studies of this kind.

It should also be noted that the measure used to assess knowledge about crystal methamphetamine was developed by the research team in response to a lack of published measures for this purpose. As such, the psychometric properties of the measure are unknown, including whether knowledge as measured in this study is a multidimensional or unidimensional construct. Therefore, direct conclusions about improving which aspects of knowledge (or knowledge in general) might reduce stigmatizing attitudes cannot be made.

Conclusions

This study is one of the most comprehensive and rigorous evaluations of a centralized national eHealth resource for the general public, focusing on substance use. CITI, an innovative portal for evidence-based information and resources about crystal methamphetamine, was found to achieve its aim of disseminating evidence-based, nonstigmatizing, and useful information and resources about crystal methamphetamine to key end user groups. Furthermore, it received good usability scores across its target groups and was found to be easy to use. In addition, interaction with CITI led to immediate improvements in knowledge and a decrease in stigmatizing attitudes about the drug, which is promising given the links among knowledge, stigma, and help-seeking. CITI demonstrates the important role of digital information and support platforms for translating evidence into practice as well as improving knowledge and reducing stigma around crystal methamphetamine. It is hoped that this will ultimately lead to an increase in help-seeking and evidence-based service use.

Acknowledgments

CITI is supported by funding from the Australian Government Department of Health. This study was also supported by the Australian National Health and Medical Research Council via Fellowships (NN, APP1166377; LS, APP1132853; KC, APP1120641; FKL, GNT1110371; and MT, APP1078407) and the Centre of Research Excellence in the Prevention and Early Intervention in Mental Illness and Substance Use (APP11349009). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors would like to thank the individuals who generously shared their experiences with them and allowed them to use their stories on the CITI website and many community members from around the country who provided their input and feedback during the development and evaluation of this resource. CITI was developed by researchers at the Matilda Centre for Research in Mental Health and Substance Use at the University of Sydney (SK, LB, HD, NN, LS, KC, FKL, MT, and CC). The authors received no financial or other benefits from the CITI website.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Ratings on Likert scale items about the purpose, goals, and evidence base of the website.
[DOCX File , 21 KB - jmir_v23i8e29026_app1.docx]

Multimedia Appendix 2
Examples of open-ended comments from the study participants.
[DOCX File , 20 KB - jmir_v23i8e29026_app2.docx]

References

1. Cho AK. Ice: a new dosage form of an old drug. Science 1990 Aug 10;249(4969):631-634. [doi: 10.1126/science.249.4969.631] [Medline: 17831955]
2. McKetin R, Kelly E, McLaren J. The relationship between crystalline methamphetamine use and methamphetamine dependence. Drug Alcohol Depend 2006 Dec 1;85(3):198-204. [doi: 10.1016/j.drugalcdep.2006.04.007] [Medline: 16723192]
3. Degenhardt L, Sara G, McKetin R, Roxburgh A, Dobbins T, Farrell M, et al. Crystalline methamphetamine use and methamphetamine-related harms in Australia. Drug Alcohol Rev 2017 Mar 11;36(2):160-170. [doi: 10.1111/dar.12426] [Medline: 27286742]
4. Australian Institute of Health and Welfare. National Drug Strategy Household Survey 2019. Canberra: Australian Institute of Health and Welfare; 2020.
5. Australian Institute of Health and Welfare. National Drug Strategy Household Survey 2016: Detailed Findings. Canberra: Australian Institute of Health and Welfare; 2017.
6. Cohn A, O'Connor R, Lancaster K, Rawstorne P, Nathan S. Media and political framing of crystal methamphetamine use in Australia. Drugs: Educ Preven Pol 2019 Oct 22;27(4):261-270. [doi: 10.1080/09687637.2019.1679089]
7. Cracks in the Ice Online Community Toolkit. The Matilda Centre for Research in Mental Health and Substance Use. URL: https://www.cracksintheice.org.au [accessed 2021-07-16]
8. Champion KE, Chapman C, Newton NC, Brierley M, Stapinski L, Kay-Lambkin F, et al. A web-based toolkit to provide evidence-based resources about crystal methamphetamine for the Australian community: collaborative development of cracks in the ice. JMIR Ment Health 2018 Mar 20;5(1):e21. [doi: 10.2196/mental.8891] [Medline: 29559427]

9. Birrell L, Deen H, Champion KE, Newton NC, Stapinski LA, Kay-Lambkin F, et al. A mobile app to provide evidence-based information about crystal methamphetamine (ice) to the community (cracks in the ice): co-design and beta testing. JMIR Mhealth Uhealth 2018 Dec 20;6(12):e11107 [FREE Full text] [doi: 10.2196/11107] [Medline: 30573443]
10. Roche A, McEntee A. Ice and the outback: Patterns and prevalence of methamphetamine use in rural Australia. Aust J Rural Health 2017 Aug 21;25(4):200-209. [doi: 10.1111/ajr.12331] [Medline: 27868256]
11. National Wastewater Drug Monitoring Program Report 11. In: Australian Criminal Intelligence Commission. Canberra: ACIC; 2020.
12. Thomas SL, Wakerman J, Humphreys JS. Ensuring equity of access to primary health care in rural and remote Australia - what core services should be locally available? Int J Equity Health 2015 Oct 29;14(1):111 [FREE Full text] [doi: 10.1186/s12939-015-0228-1] [Medline: 26510998]
13. Bennett GG, Glasgow RE. The delivery of public health interventions via the internet: actualizing their potential. Annu Rev Public Health 2009 Apr;30(1):273-292. [doi: 10.1146/annurev.publhealth.031308.100235] [Medline: 19296777]
14. Fox MP. A systematic review of the literature reporting on studies that examined the impact of interactive, computer-based patient education programs. Patient Educ Couns 2009 Oct;77(1):6-13. [doi: 10.1016/j.pec.2009.02.011] [Medline: 19345550]
15. Woolf SH. The meaning of translational research and why it matters. J Am Med Assoc 2008 Jan 9;299(2):211-213. [doi: 10.1001/jama.2007.26] [Medline: 18182604]
16. Schneider F, van Osch L, de Vries H. Identifying factors for optimal development of health-related websites: a delphi study among experts and potential future users. J Med Internet Res 2012 Feb 14;14(1):e18 [FREE Full text] [doi: 10.2196/jmir.1863] [Medline: 22357411]
17. de Argaez E. Internet Users and 2021 Population in Oceania. Internet World Stats Usage and Population Statistics. URL: https://www.internetworldstats.com/stats6.htm [accessed 2020-10-28]
18. Ryan GS, Haroon M, Melvin G. Evaluation of an educational website for parents of children with ADHD. Int J Med Inform 2015 Nov;84(11):974-981. [doi: 10.1016/j.ijmedinf.2015.07.008] [Medline: 26265480]
19. 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. J Am Med Assoc 2002 May 22;287(20):2691-2700. [doi: 10.1001/jama.287.20.2691] [Medline: 12020305]
20. Godinho MA, Ansari S, Guo GN, Liaw ST. Toolkits for implementing and evaluating digital health: a systematic review of rigor and reporting. J Am Med Inform Assoc 2021 Jun 12;28(6):1298-1307. [doi: 10.1093/jamia/ocab010] [Medline: 33619519]
21. Deen H, Kershaw S, Newton N, Stapinski L, Birrell L, Debenham J, et al. Stigma, discrimination and crystal methamphetamine ('ice'): current attitudes in Australia. Int J Drug Policy 2021 Jan;87:102982. [doi: 10.1016/j.drugpo.2020.102982] [Medline: 33126164]
22. Regen M. Drug users' lay consultation processes: symptom identification and management. Adc Med Soc 2002;8:323-341. [doi: 10.1016/s1057-6290(02)80032-8]
23. Cama E, Brener L, Wilson H, von Hippel C. Internalized stigma among people who inject drugs. Subst Use Misuse 2016 Oct 14;51(12):1664-1668. [doi: 10.1080/10826084.2016.1188951] [Medline: 27459264]
24. Luoma JB, Twohig MP, Waltz T, Hayes SC, Roget N, Padilla M, et al. An investigation of stigma in individuals receiving treatment for substance abuse. Addict Behav 2007 Jul;32(7):1331-1346. [doi: 10.1016/j.addbeh.2006.09.008] [Medline: 17092656]
25. Brooke J. SUS - a quickdirty usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation In Industry. Boca Raton, Florida: CRC Press; 1996:189-193.
26. Sauro J. Measuring usability with the System Usability Scale (SUS). MeasuringU. 2011 Feb 3. URL: http://www.measuringu.com/sus.php [accessed 2021-02-05]
27. Brener L, von Hippel W. Measuring attitudes toward injecting drug users and people with hepatitis C. Subst Use Misuse 2008 Jul 3;43(3-4):295-302. [doi: 10.1080/10826080701202627] [Medline: 18365932]
28. Griffiths KM, Christensen H, Jorm AF, Evans K, Groves C. Effect of web-based depression literacy and cognitive-behavioural therapy interventions on stigmatising attitudes to depression: randomised controlled trial. Br J Psychiatry 2004 Oct 2;185(4):342-349. [doi: 10.1192/bjp.185.4.342] [Medline: 15458995]
29. Griffiths KM, Nakane Y, Christensen H, Yoshioka K, Jorm AF, Nakane H. Stigma in response to mental disorders: a comparison of Australia and Japan. BMC Psychiatry 2006 May 23;6(1):1-12. [doi: 10.1186/1471-244x-6-21]
30. Yap MB, Mackinnon A, Reavley N, Jorm AF. The measurement properties of stigmatizing attitudes towards mental disorders: results from two community surveys. Int J Methods Psychiatr Res 2014 Mar 18;23(1):49-61 [FREE Full text] [doi: 10.1002/mpr.1433] [Medline: 24550065]
31. 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]
32. Australian Demographic Statistics, Jun 2018. Australian Bureau of Statistics. URL: https://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/3101.0Main+Features1Jun%202018?OpenDocument= [accessed 2021-07-16]
33. Sprenkle B. Performance Benchmarks for Healthcare SEO. 9 Clouds. URL: https://9clouds.com/blog/performance-benchmarks-for-healthcare-seo/ [accessed 2021-02-01]

34. Kuek A, Hakkennes S. Healthcare staff digital literacy levels and their attitudes towards information systems. Health Informatics J 2020 Mar 15;26(1):592-612 [FREE Full text] [doi: 10.1177/1460458219839613] [Medline: 30983476]
35. Milin R, Kutcher S, Lewis SP, Walker S, Wei Y, Ferrill N, et al. Impact of a mental health curriculum on knowledge and stigma among high school students: a randomized controlled trial. J Am Acad Child Adolesc Psychiatry 2016 May;55(5):383-91.e1. [doi: 10.1016/j.jaac.2016.02.018] [Medline: 27126852]
36. Jorm AF, Wright A. Influences on young people's stigmatising attitudes towards peers with mental disorders: national survey of young Australians and their parents. Br J Psychiatry 2008 Feb 2;192(2):144-149. [doi: 10.1192/bjp.bp.107.039404] [Medline: 18245033]
37. Simmons L, Jones T, Bradley E. Reducing mental health stigma: the relationship between knowledge and attitude change. EJMH 2017;12(1):25-40. [doi: 10.5708/ejmh.12.2017.1.2]
38. Smith V, Reddy J, Foster K, Asbury ET, Brooks J. Public perceptions, knowledge and stigma towards people with schizophrenia. J of Public Mental Health 2011 Mar 15;10(1):45-56. [doi: 10.1108/17465721111134547]
39. Cumming C, Troeung L, Young JT, Kelty E, Preen DB. Barriers to accessing methamphetamine treatment: a systematic review and meta-analysis. Drug Alcohol Depend 2016 Nov 1;168:263-273. [doi: 10.1016/j.drugalcdep.2016.10.001] [Medline: 27736680]
40. Jeong D, Cheng M, St-Jean M, Jalali A. Evaluation of ementalhealth.Ca, a Canadian mental health website portal: mixed methods assessment. JMIR Ment Health 2019 Sep 6;6(9):e13639 [FREE Full text] [doi: 10.2196/13639] [Medline: 31493328]
41. Peacock A, Uporova J, Karlsson A, Price O, Gibbs D, Swanton R, et al. Australian Drug Trends 2020: Key Findings from the Illicit Drug Reporting System (IDRS) Interviews. Sydney, Australia: National Drug and Alcohol Research Centre, UNSW Sydney; 2021.
42. Pang PC, Chang S, Verspoor K, Pearce J. Designing health websites based on users' web-based information-seeking behaviors: a mixed-method observational study. J Med Internet Res 2016 Jun 6;18(6):e145 [FREE Full text] [doi: 10.2196/jmir.5661] [Medline: 27267955]
43. Driesenaar JA, van Dulmen S, van Weert JC, Noordman J. Patients’ evaluation of a preparatory online communication tool for older patients with cancer preceding chemotherapy. Cancer Nurs 2018 Nov 15;43(2):E71-E78. [doi: 10.1097/ncc.0000000000000660]
44. Barbara AM, Dobbins M, Haynes RB, Iorio A, Lavis JN, Raina P, et al. The mcmaster optimal aging portal: usability evaluation of a unique evidence-based health information website. JMIR Hum Factors 2016 May 11;3(1):e14 [FREE Full text] [doi: 10.2196/humanfactors.4800] [Medline: 27170443]
45. Knijnenburg SL, Kremer LC, Versluys AB, Braam KI, Mud MS, van der Pal HJ, et al. Evaluation of a patient information website for childhood cancer survivors. Support Care Cancer 2013 Apr 25;21(4):919-926. [doi: 10.1007/s00520-012-1604-7] [Medline: 23007883]
46. Janssen A, Robinson TE, Provan P, Shaw T. The sydney west knowledge portal: evaluating the growth of a knowledge portal to support translational research. J Med Internet Res 2016 Jun 29;18(6):e170 [FREE Full text] [doi: 10.2196/jmir.5786] [Medline: 27357641]
47. Young M, Richards C, Gunning M. Online mental health resources for teenagers: an evaluation of two websites developed for adolescents. Adv School Mental Health Prom 2012 Oct;5(4):277-289. [doi: 10.1080/1754730x.2012.736792]
48. Buteau-Poulin A, Gosselin C, Bergeron-Ouellet A, Kiss J, Lamontagne M, Maltais D, et al. Availability and quality of web resources for parents of children with disability: content analysis and usability study. JMIR Pediatr Parent 2020 Nov 10;3(2):e19669 [FREE Full text] [doi: 10.2196/19669] [Medline: 33170127]
49. Cummins CO, Prochaska JO, Driskell M, Evers KE, Wright JA, Prochaska JM, et al. Development of review criteria to evaluate health behavior change websites. J Health Psychol 2003 Jan 1;8(1):55-62. [doi: 10.1177/1359105303008001434] [Medline: 22113900]
50. Enggasser JL, Livingston NA, Ameral V, Brief DJ, Rubin A, Helmuth E, et al. Public implementation of a web-based program for veterans with risky alcohol use and PTSD: A RE-AIM evaluation of VetChange. J Subst Abuse Treat 2021 Mar;122:108242. [doi: 10.1016/j.jsat.2020.108242] [Medline: 33509419]
51. 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 Jun 26;22(6):e16701 [FREE Full text] [doi: 10.2196/16701] [Medline: 32442134]

Abbreviations

CITI: Cracks in the Ice
DSS: Depression Stigma Scale
SUS: System Usability Scale

Edited by R Kukafka; submitted 23.03.21; peer-reviewed by P Atorkey, J Dol; comments to author 14.05.21; revised version received 30.05.21; accepted 28.06.21; published 10.08.21.

Please cite as:
Kershaw S, Birrell L, Deen H, Newton NC, Stapinski LA, Champion KE, Kay-Lambkin F, Teesson M, Chapman C
Evaluation of a Digital Health Initiative in Illicit Substance Use: Cross-sectional Survey Study
J Med Internet Res 2021;23(8):e29026
URL: https://www.jmir.org/2021/8/e29026
doi:10.2196/29026
PMID:34383690

©Steph Kershaw, Louise Birrell, Hannah Deen, Nicola C Newton, Lexine A Stapinski, Katrina E Champion, Frances Kay-Lambkin, Maree Teesson, Cath Chapman. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Effectiveness of Internet-Based Cognitive Behavior Therapy (Fatigue in Teenagers on the Internet) for Adolescents With Chronic Fatigue Syndrome in Routine Clinical Care: Observational Study

Eline Albers[1*], MSc; Linde N Nijhof[1*], MSc; Emma E Berkelbach van der Sprenkel[1], MSc; Elise M van de Putte[1], MD, PhD; Sanne L Nijhof[1], MD, PhD; Hans Knoop[2], PhD

[1]Department of Pediatrics, Wilhelmina Children's Hospital, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands

[2]Department of Medical Psychology, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam Public Health Research Institute, Amsterdam, Netherlands

[*]these authors contributed equally

Corresponding Author:
Hans Knoop, PhD
Department of Medical Psychology
Amsterdam University Medical Centers, University of Amsterdam
Amsterdam Public Health Research Institute
Meibergdreef 9
Amsterdam, 1105 AZ
Netherlands
Phone: 31 20 566 4661
Email: hans.knoop@amsterdamumc.nl

Abstract

Background: Internet-based cognitive behavior therapy (I-CBT) for adolescents with chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME) has been shown to be effective in a randomized controlled trial (RCT; Fatigue in Teenagers on the Internet [FITNET]). FITNET can cause a significant reduction in fatigue and disability.

Objective: We aimed to investigate whether FITNET treatment implemented in routine clinical care (IMP-FITNET) was as effective, using the outcomes of the FITNET RCT as the benchmark.

Methods: Outcomes of CFS/ME adolescents who started IMP-FITNET between October 2012 and March 2018 as part of routine clinical care were compared to the outcomes in the FITNET RCT. The primary outcome was fatigue severity assessed posttreatment. The secondary outcomes were self-reported physical functioning, school attendance, and recovery rates. Clinically relevant deterioration was assessed posttreatment, and for this outcome, a face-to-face CBT trial was used as the benchmark. The attitude of therapists toward the usability of IMP-FITNET was assessed through semistructured interviews. The number of face-to-face consultations during IMP-FITNET was registered.

Results: Of the 384 referred adolescents with CFS/ME, 244 (63.5%) started IMP-FITNET, 84 (21.9%) started face-to-face CBT, and 56 (14.6%) were not eligible for CBT. Posttreatment scores for fatigue severity (mean 26.0, SD 13.8), physical functioning (mean 88.2, SD 15.0), and full school attendance (mean 84.3, SD 26.5) fell within the 95% CIs of the FITNET RCT. Deterioration of fatigue and physical functioning after IMP-FITNET was observed at rates of 1.2% (n=3) and 4.1% (n=10), respectively, which is comparable to a waiting list condition (fatigue: 1.2% vs 5.7%, $\chi^2_1=3.5$, *P*=.06; physical functioning: 4.1% vs 11.4%, $\chi^2_1=3.3$, *P*=.07). Moreover, 41 (16.8%) IMP-FITNET patients made use of face-to-face consultations.

Conclusions: IMP-FITNET is an effective and safe treatment for adolescents with CFS/ME in routine clinical care.

(J Med Internet Res 2021;23(8):e24839) doi:10.2196/24839

KEYWORDS

Fatigue in Teenagers on the Internet; cognitive behavior therapy; fatigue; chronic fatigue syndrome; adolescents; implementation

Introduction

Chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME) is a disabling condition in which patients have severe, medically unexplained, and persistent (>6 months) fatigue, resulting in impairment of functioning [1]. The prevalence of CFS/ME in adolescents is 0.11% to 1.29% in Dutch and British adolescent populations, with a female-to-male ratio of 2:1 to 5:1 [2,3]. In adolescents, CFS/ME often has a chronic course, leading to school absence, and has long-term detrimental effects on social and academic development [4,5].

The etiology of CFS/ME is unknown. The context of the biopsychosocial model defines individual predisposing, precipitating, and perpetuating factors that provoke and maintain severe fatigue and disability [6]. Face-to-face cognitive behavior therapy (CBT) aimed at fatigue-maintaining factors has been tested in several randomized controlled trials (RCTs) and leads to a significant reduction in fatigue and an improvement in physical functioning and school attendance [4,7,8]. An internet-based format of CBT for adolescents with CFS/ME, named Fatigue in Teenagers on the Internet (FITNET), was found to be effective in an RCT, leading to significant reduction of fatigue and fatigue-related disabilities [4]. Two-thirds of patients reported fatigue levels and physical functioning within the normal limits, as well as full school attendance following treatment [4]. Currently, in the United Kingdom, an RCT is investigating the feasibility, clinical effectiveness, and cost-effectiveness of FITNET delivered in the context of the National Health Service (FITNET_NHS) [9,10]. FITNET is easily accessible since it is not bound to the geographic location of the therapist delivering the intervention [4]. Patients do not need to travel to a treatment center and can follow the treatment at home, making the intervention easy to follow. Importantly, adolescents with CFS preferred FITNET over face-to-face treatment [4]. It is not self-evident that outcomes of an RCT can be extrapolated to routine clinical care (RCC), since the effectiveness of an RCT may be overestimated due to strict inclusion criteria and close monitoring of the intervention [11,12]. Thus far, the effectiveness of FITNET delivered in RCC has not been shown. Recently, concerns have been raised about the safety of behavioral interventions for CFS/ME. It has been suggested that CBT leads to deterioration of symptoms and functioning [13]. Analyses of RCTs testing the efficacy of CBT for CFS have thus far not shown a higher prevalence of deterioration or more adverse events following or during CBT compared to control conditions [4,14,15]. However, this has not yet been determined for internet-based cognitive behavior therapy (I-CBT) delivered in RCC for adolescents, which uses less strict inclusion criteria and less stringent monitoring of the treatment process.

The primary aim of this study was to determine whether FITNET implemented in RCC (IMP-FITNET) is as effective as in a research context with respect to the outcomes of fatigue severity, physical functioning, school attendance, and recovery rates, using the outcomes of the previous RCT as the benchmark [4]. The secondary aim was to investigate the safety of IMP-FITNET by assessing the frequency of deterioration of fatigue and physical functioning in comparison to a waiting list condition of another benchmark RCT in adolescents [14]. Lastly, the attitude of therapists toward the usability of IMP-FITNET was assessed through semistructured interviews.

Methods

Design and Participants

This was an observational study of RCC. Data were collected retrospectively from adolescents who finished treatment in RCC after implementation.

All patients were referred to the Expert Centre for Chronic Fatigue (ECCF), a national referral center for IMP-FITNET and face-to-face CBT for adolescents with CFS/ME, and were retrospectively included in the study between October 2012 and March 2018.

The inclusion criteria were as follows: (1) CFS according to the US Centers for Disease Control (CDC) criteria revised in 2003 [1,16]; (2) 12-18 years of age at baseline; (3) severe fatigue, operationalized as a score of 40 or higher on the fatigue severity subscale of the Checklist Individual Strength-20 (CIS-20) [17]; (4) self-reported substantial disabilities in daily functioning; (5) access to the internet; and (6) no psychiatric comorbidity that could explain the presence of fatigue, ruled out by the Mini International Neuropsychiatric Interview for children (M.I.N.I. KID) [18].

Posttreatment effectiveness in terms of fatigue severity, physical functioning, school attendance, and recovery rates were compared with results derived from the previously published RCT on FITNET [4]. The safety of IMP-FITNET was determined by comparing the prevalence of the deterioration of fatigue and physical functioning with the results of a benchmark study that reported on deterioration rates in a waitlist condition of an RCT testing the efficacy of face-to-face CBT for CFS/ME in adolescents [14]. In RCC, patients received either IMP-FITNET or face-to-face CBT as decided in a shared decision process, in which patients/parents and therapists worked together to choose the best suitable therapy, reflecting both evidence and patient priorities and preferences.

The Dutch Medical Research Involving Human Subjects Act did not apply to our study, as the collected data were part of RCC. Therefore, no formal ethical approval from the medical ethics committee was needed for this study.

Treatment

CBT for CFS/ME is developed on the basis of a cognitive behavioral model of CFS, assuming that behavior and beliefs can perpetuate symptoms [19]. In this study, two formats of CBT for adolescents with CFS/ME were used. A face-to-face CBT treatment manual for adolescents was applied, which was found to be efficacious in previous research [7]. The second format was an implemented version of the online CBT intervention FITNET with the same content and similar layout, but using a different software package, referred to as IMP-FITNET. FITNET is an internet-based CBT program developed on the basis of the face-to-face CBT treatment manual for adolescents. FITNET has been found to be efficacious in the context of research [4]. The FITNET program consists, aside from psychoeducation on CFS, of 21 interactive treatment modules, with content in line with face-to-face CBT treatment for adolescents and a program for care givers. Each module includes assignments and mandatory e-consultations with the therapist [4,9,20]. IMP-FITNET has the same 21 interactive modules for patients with e-consultations and caregivers. In IMP-FITNET, therapists were however allowed to offer face-to-face consultations or phone calls if deemed necessary, and for patients starting treatment in 2017, videoconferencing was possible.

Eleven trained cognitive behavioral therapists who received weekly supervision from experienced clinical psychologists delivered face-to-face CBT and IMP-FITNET.

Outcome Variables

Primary Outcome Posttreatment

All outcome variables were self-reported. The primary outcome was fatigue severity assessed with the subscale fatigue severity of the CIS-20. This subscale consists of eight items scored on a 7-point Likert scale, resulting in a fatigue severity score ranging from 8 to 56. A score ≥40 indicates the presence of severe fatigue [17]. The internal consistency and discriminative validity of the CIS are excellent [7,17].

Secondary Outcomes Posttreatment

Physical functioning was measured with the subscale physical functioning (nine items, range 0%-100%) of the Child Health Questionnaire-87 (CHQ-87). The questionnaire is validated and has good internal consistency [21].

School presence was assessed using a diary and reported as the percentage of classes attended over the past 2 weeks divided by the scheduled number of classes for peers [4].

Recovery was defined in relation to healthy peers by having a CIS-fatigue score <40 [2], a CHQ-87 physical score ≥85% [21], and school absence ≤10% in the past 2 weeks [2]. These recovery criteria were derived from the FITNET RCT [4].

Deterioration of fatigue was defined as an increase of more than six points in CIS-fatigue, and deterioration of physical functioning was defined as a decrease of more than 10 points in CHQ-physical [9,22].

The benchmark for deterioration, as a proxy for safety, was the waiting list condition in a prior RCT on the efficacy of face-to-face CBT [14]. Instead of the CHQ-87 subscale physical functioning, the SF-36 (short-form) of the RAND was used in this study [23].

Semistructured Interviews

Using a semistructured telephone interview, 11 therapists were asked which criteria they used to propose to start with either IMP-FITNET or face-to-face CBT for the individual adolescent or face-to-face consultations during IMP-FITNET. Nine therapists participated. Interviews with the therapist were recorded and transcribed by one researcher (EA). The themes were independently synthesized by two researchers (EA and LNN) based on the interviews [24]. Discrepancies were resolved through discussion with the principal investigator (HK) to reach consensus. In addition, the number of face-to-face consultations during IMP-FITNET was registered.

Procedure

After referral, adolescents had two diagnostic face-to-face sessions with a psychologist. The results of the baseline assessment were discussed with the adolescent and parents, and this was followed by a shared decision for either IMP-FITNET or face-to-face CBT. Following treatment, adolescents completed an online posttreatment assessment, which was discussed in a face-to-face session.

Statistical Analysis

The demographic characteristics of the adolescents and baseline scores were compared using a benchmark strategy, in which the baseline scores of RCC were compared with the 95% CIs of corresponding values in the FITNET RCT. If the mean value in RCC was outside the 95% CI of the RCT, it was considered divergent. The same procedure was used to compare the post-treatment outcomes of RCC with the RCT. Baseline characteristics of patients lost to follow-up were compared with those who were assessed posttreatment using a t test for independent groups.

Analyses were based on intention to treat, using the summary estimate of five imputations for 15 missing observations in the primary outcome, with the assumption that data were missing at random [25]. All baseline and posttreatment scores were entered as predictors. The posttreatment score of fatigue severity was entered as a variable to impute. Moreover, analyses were repeated with only those patients who met all the inclusion criteria of the benchmark FITNET RCT (baseline score for physical activity <85% and/or school presence ≤85%). The within-group Cohen d was reported as the effect size. For the main outcome fatigue severity, we calculated the percentage of patients who scored below the cutoff of severe fatigue (CIS <40) and reported a reliable change index (z) score greater than +1.96 [26]. To test differences in deterioration, chi-square tests were performed. Percentages of adolescents who were lost to follow-up in the RCT and in RCC were compared with chi-square tests. Lastly, the number of face-to-face consultations during IMP-FITNET were registered.

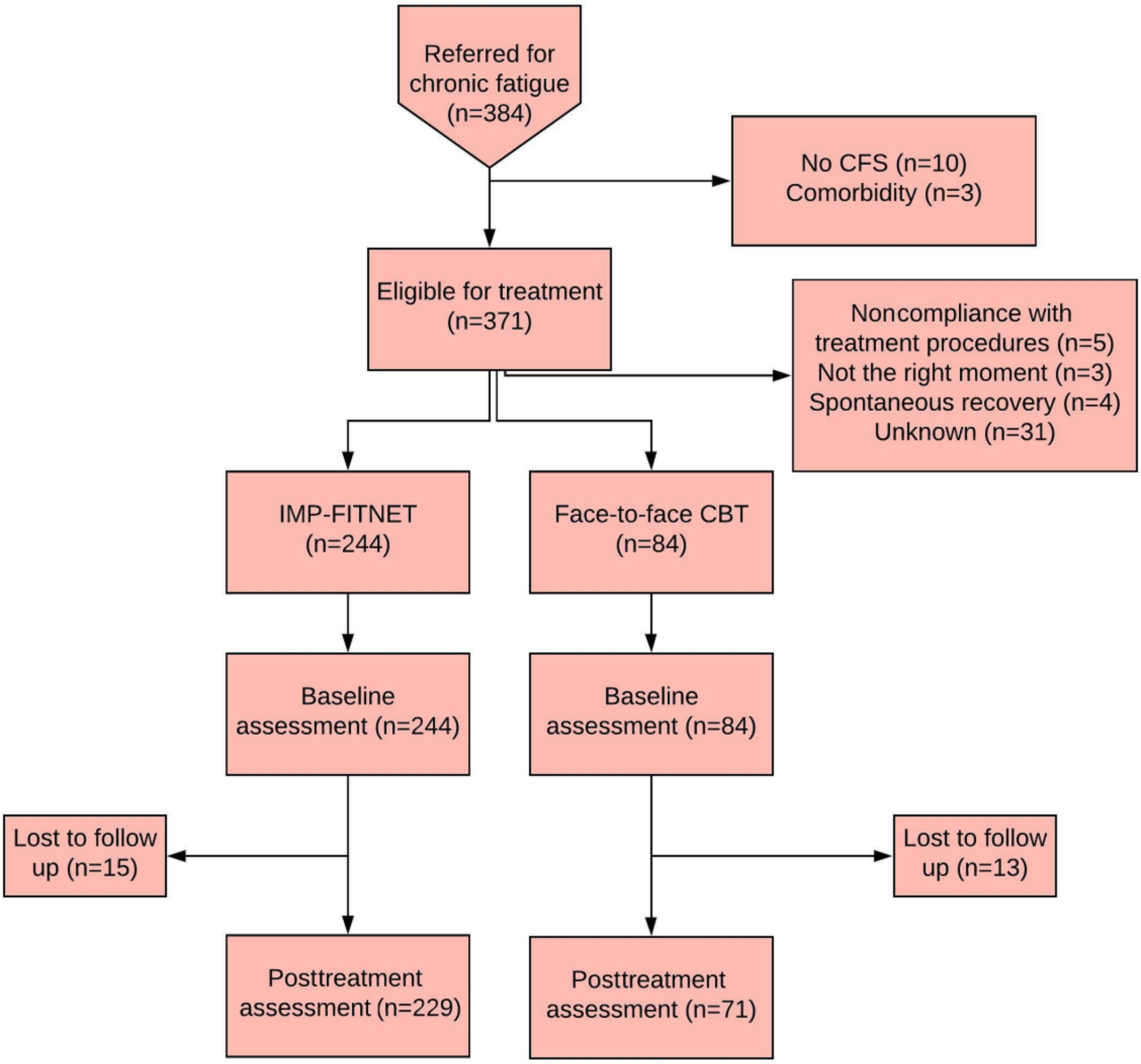
SPSS version 25 (IBM Corp) was used for statistical analyses, and significance was set at $P<.05$.

Results

Study Population

Of the 384 referred adolescents, 371 were eligible for treatment, of which 328 (88.4%) started treatment. Of the 328 patients, 244 (74.4%) received IMP-FITNET and 84 (25.6%) received face-to-face CBT. All 328 adolescents filled out the baseline assessment, 229 of the 244 patients (93.8%) who received IMP-FITNET completed the posttreatment assessment, and 71 of the 84 patients (84.5%) who received face-to-face CBT completed the posttreatment assessment (Figure 1).

Figure 1. Study flow of the patients in routine clinical care. CBT: cognitive behavioral therapy; CFS: chronic fatigue syndrome; IMP-FITNET: implemented Fatigue in Teenagers on the Internet.


Referred for chronic fatigue (n=384)
No CFS (n=10)
Comorbidity (n=3)
Eligible for treatment (n=371)
Noncompliance with treatment procedures (n=5)
Not the right moment (n=3)
Spontaneous recovery (n=4)
Unknown (n=31)
IMP-FITNET (n=244)
Face-to-face CBT (n=84)
Baseline assessment (n=244)
Baseline assessment (n=84)
Lost to follow up (n=15)
Lost to follow up (n=13)
Posttreatment assessment (n=229)
Posttreatment assessment (n=71)

Baseline Characteristics

Adolescents lost to follow-up differed significantly in physical functioning at baseline (mean score 64.7, SD 16.4 vs mean score [lost to follow-up] 73.0, SD 19.4; t_{241}=−2.04, *P*=.04).

Table 1 provides the baseline characteristics of the adolescents in RCC (IMP-FITNET and face-to-face CBT) compared with the baseline characteristics of the patients from the benchmark FITNET RCT. The duration of symptoms was significantly lower in patients who received IMP-FITNET and significantly higher in patients who received face-to-face CBT compared to the FITNET RCT. Both were outside the 95% CI of the FITNET RCT. The duration of symptoms between IMP-FITNET and face-to-face CBT did not significantly differ (t_{306}=0.984, *P*=.33). Fatigue severity of the adolescents following either IMP-FITNET or face-to-face CBT was lower and their physical functioning was higher than in the benchmark study, as was their school participation. Moreover, adolescents receiving face-to-face CBT were younger and more often male than in the benchmark study. The analyses were repeated with the subset of patients in RCC who met all the inclusion criteria of the benchmark study. This analysis showed the same pattern of results (data not shown).

The percentage of adolescents lost to follow-up was significantly higher in IMP-FITNET than in the FITNET RCT (8.5% [n=28] vs 3.0% [n=4], N=463, χ^2_1=3.3, *P*=.03).

Table 1. Baseline characteristics of the patients in routine clinical care and the 95% CIs of the benchmark Fatigue in Teenagers on the Internet randomized controlled trial scores.

Variable	IMP-FITNET[a] (N=244)	F2F-CBT[b] (N=84)	95% CI benchmark FITNET RCT[c] (N=135)[d]
Age at entry (years), mean (SD)	16.1 (1.4)	15.4 (1.8)	15.6-16.1[e]
Gender (female), n (%)	202 (82.8%)	62 (73.8%)	76%-89%[e]
Duration of symptoms at entry (months), median (range)	18 (3-96)	30 (6-96)	20-27 months[e]
Fatigue severity (CIS[f]), mean (SD)	49.8 (5.0)	48.7 (6.3)	50.6-52.2[e]
School attendance, mean % (SD)	60.5 (34.2)	61.1 (33.4)	37.0-47.6[e]
Number of children with >85% school attendance, n/N (%)	70/244 (31.3%)	22/84 (28.9%)	5%-15%[e]
Physical functioning (CHQ-87[g]), mean (SD)	64.7 (16.4)	63.2 (19.3)	55.7-61.8[e]
Anxiety score (STAIC[h]), mean (SD)	31.9 (7.4)	34.7 (8.0)	31.9-34.4[e]

[a] IMP-FITNET: implemented Fatigue in Teenagers on the Internet.

[b] F2F-CBT: face-to-face cognitive behavior therapy.

[c] RCT: randomized controlled trial.

[d] Benchmark FITNET RCT: the study by Nijhof et al [4].

[e] 95% CIs of the values of the benchmark FITNET RCT.

[f] CIS: Checklist Individual Strength.

[g] CHQ-87: Child Health Questionnaire-87.

[h] STAIC: State-Trait Anxiety Inventory for Children.

Primary and Secondary Outcomes Posttreatment

The CIS-fatigue severity score after IMP-FITNET and face-to-face CBT in RCC fell within the 95% CI of the benchmark FITNET study. Additionally, 173 of the 229 adolescents (75.5%) with a posttreatment fatigue score had a reliable change index score greater than +1.96 and a score lower than 40 on the CIS. All secondary outcomes (Table 2) fell within the 95% CIs of the benchmark study. The analyses were repeated with the subset of patients in RCC who fulfilled all the inclusion criteria of the benchmark study. This analysis showed the same pattern of results (data shown in Multimedia Appendix 1).

In RCC, 3 of the 244 patients (1.2%) reported a clinically significant deterioration of fatigue severity after IMP-FITNET. In the waiting list condition of a face-to-face CBT benchmark study, 2 out of 35 patients (5.7%) showed clinically significant deterioration in fatigue severity [15]. This did not significantly differ (N=279; χ^2_1=3.5, *P*=.06). Nine of the 244 patients (3.7%) following IMP-FITNET had a baseline score for fatigue severity above 50 and did not show improvement. They could not be identified as patients showing clinically significant deterioration of fatigue due to the ceiling effect of the CIS-20 questionnaire (maximum score of 56).

In RCC, 10 of the 244 patients (4.1%) reported clinically significant deterioration of physical functioning after IMP-FITNET. In the benchmark study [14], 4 of 35 patients (11.4%) showed clinically significant deterioration of physical functioning. Deterioration of physical functioning did not significantly differ between the waiting list condition of the benchmark and IMP-FITNET (N=279; χ^2_1=3.3, *P*=.07) (Table 3).

The within-treatment group effect size of FITNET in the RCT was large (Cohen *d*=2.73), with a 95% CI of 2.26 to 3.21. The effect size of IMP-FITNET was also large (Cohen *d*=2.28) and fell within the 95% CI of the FITNET RCT.

Table 2. Posttreatment scores of patients in routine clinical care and the 95% CIs of the benchmark Fatigue in Teenagers on the Internet randomized controlled trial scores.

Variable	IMP-FITNET[a] (N=229)	F2F-CBT[b] (N=71)	95% CI benchmark FITNET RCT[c] (N=67)[d]
Fatigue severity (CIS[e]), mean (SD)	26.0 (13.8)	25.8 (12.3)	20.7-27.3[f]
Physical functioning (CHQ-87[g]), mean (SD)	88.2 (15.0)	89.3 (12.8)	85.2-91.9[f]
School attendance, mean % (SD)	84.3 (26.5)	87.1 (23.6)	77.1-91.5[f]
Recovery[h], %	58%	60%	54-77%[f]

[a] IMP-FITNET: implemented Fatigue in Teenagers on the Internet.

[b] F2F-CBT: face-to-face cognitive behavior therapy.

[c] RCT: randomized controlled trial.

[d] Benchmark FITNET RCT: the study by Nijhof et al [4].

[e] CIS: Checklist Individual Strength.

[f] 95% CIs of the values of the benchmark FITNET RCT.

[g] CHQ-87: Child Health Questionnaire-87.

[h] Cutoff scores for recovery are as follows: fatigue severity of <40 on the CIS-20 subscale fatigue; school absence of ≤10%, and a physical functioning score of ≥85% on the CHQ-87 subscale physical functioning.

Table 3. Number of patients with symptom deterioration between preassessment and postassessment.

Variable	IMP-FITNET[a] (N=244)	F2F-CBT[b] (N=71)	Waiting list condition[c] (N=35)
Deterioration of fatigue severity[d], n (%)	3 (1.2%)	2 (2.8%)	2 (5.7%)
Deterioration of physical functioning[e,f], n (%)	10 (4.1%)	2 (2.8%)	4 (11.4%)

[a] IMP-FITNET: implemented Fatigue in Teenagers on the Internet.

[b] F2F-CBT: face-to-face cognitive behavior therapy.

[c] Data from adolescents on a waiting list condition in a study by Stulemeijer et al [7].

[d] Increase of >6 points on the Checklist Individual Strength (CIS).

[e] Decrease of >10 points on the Child Health Questionnaire-87 (CHQ-87) for patients following IMP-FITNET or F2F-CBT.

[f] Decrease of >10 points on the Short Form-36 (SF-36).

Semistructured Interviews and Number of Face-to-Face Consultations

Nine of the 11 therapists were interviewed. Face-to-face CBT was preferred to IMP-FITNET when there were interaction problems in the family or when the patient had psychiatric or somatic comorbidities. Therapists decided to make use of face-to-face consultations during IMP-FITNET treatment in the case of perceived inability of the patient to benefit from solely IMP-FITNET or anticipated problems with adherence and motivation. In general, therapists preferred blended therapy with combinations of IMP-FITNET.

Of the 244 adolescents who started IMP-FITNET, 116 (47.5%) followed only IMP-FITNET without face-to-face consultations and 102 (41.8%) had at least one face-to-face consultation with their therapist, and of these, 41 adolescents (16.8%) had over 3 face-to-face consultations. Adolescents who used face-to-face consultations had on average about three face-to-face consultations (mean 3.2, SD 3.81, modus 1). Moreover, videoconferencing became an additional feature during IMP-FITNET treatment for 50 patients, of which 23 patients (46.0%) used videoconferencing, with an average of 5.1 conferences (SD 3.48) lasting on average 24.3 minutes (SD 20.7).

Discussion

The posttreatment outcomes of adolescents with CFS/ME treated with I-CBT implemented in RCC (IMP-FITNET) were within the CIs of the outcomes from the benchmark with respect to levels of fatigue severity, physical functioning, school attendance, and recovery rates at posttreatment. Additionally, 133 of the 229 (58.1%) adolescents treated with IMP-FITNET met the recovery criteria posttreatment. The within-treatment group effect size of the decrease in fatigue severity with IMP-FITNET was also within the CI of the benchmark. At baseline, patients had an average fatigue severity score of 49.8 (SD 5.0), and after treatment, their fatigue severity score reduced on average by 23.8 points to 26.0 (SD 13.8). We conclude from this that IMP-FITNET applied in RCC is an effective intervention. Our findings are in line with the results of studies in adult patients with CFS/ME, in which blended CBT implemented in RCC was as effective as in a research context [27,28].

The primary and secondary outcomes of adolescents following face-to-face CBT were also within the CIs of the FITNET RCT.

A quarter of the referred patients eligible for IMP-FITNET started with face-to-face CBT after a shared decision process. This was more often the case when patients were young, were male, had a long symptom duration, and were anxious. Therapists indicated preferring face-to-face CBT in case of family interaction problems, psychiatric or somatic comorbidities, and problems with motivation.

The prevalence of a clinically significant deterioration following IMP-FITNET was low in general and comparable with a waiting list condition of a prior CBT study [14]. Therefore, we consider IMP-FITNET safe for application in RCC, without evidence of an increased risk of deterioration in fatigue and physical functioning. This replicates similar findings in previous studies of the safety of CBT for CFS/ME conducted in a research context [14].

At baseline, differences existed between adolescents who received IMP-FITNET and those in the benchmark FITNET RCT. Adolescents in RCC were less severely fatigued, were less physically impaired, had less school absence, and had a shorter symptom duration compared with patients from the benchmark RCT. This may be the result of the increased availability of an evidence-based and internet-based treatment for this patient group after nationwide implementation of I-CBT that followed the publication of the FITNET RCT results. Moreover, for these less severely affected adolescents with CFS/ME, IMP-FITNET is effective. IMP-FITNET has the advantage that adolescents do not need to travel for treatment. The nationwide availability of an effective intervention favors earlier referral.

This clinical observational study was not designed to investigate the difference in effectiveness of IMP-FITNET versus face-to-face CBT in relation to specific patient populations. Nevertheless, we found that even with a less strict treatment protocol and a more blended form of treatment, IMP-FITNET is effective. Although CBT for adolescents with CFS/ME (FITNET, face-to-face CBT, or IMP-FITNET) is considered effective, one-third of patients do not recover. To further improve the treatment and prognosis of adolescent CFS/ME, it is important to identify the factors that contribute to treatment effectiveness and assess which factors are associated with nonrecovery. Some issues need further consideration. First, in this observational study design, the choice of the treatment form (face to face vs IMP-FITNET) was determined by health care providers taking into account the patient's preference. For this reason, there are methodological limitations, and the most important one is the risk of confounding by indication [29]. Second, the inclusion criteria of the FITNET RCT were stricter than those of IMP-FITNET, applying cutoff scores for physical functioning or school participation [4]. We did not find evidence that this influenced the outcomes of IMP-FITNET as the pattern of results was similar in the subgroup of patients who followed IMP-FITNET and met the stricter inclusion criteria. Next, the average time between pretreatment and posttreatment assessments in RCC was much longer than in the FITNET RCT, owing to the waiting list, holidays, and breaks. Third, not all results of the RCT could be compared with the findings of this study since some data were obtained with a different method or were not systematically assessed in the implementation study (eg, self-reported recovery). Fourth, although still relatively low, the dropout rate in RCC was significantly higher than in the FITNET RCT. This could, despite imputation of missing data, have led to selection bias. Lastly, the benchmark used to compare deterioration of physical functioning did not use the exact same questionnaire. As this study reported on data from patients who were treated with an evidence-based treatment in RCC, we do attribute the reduction in symptoms to the treatment with IMP-FITNET. Moreover, the advantage of retrospective data from RCC is the unbiased representation of the patient population.

IMP-FITNET in RCC was adapted according to therapist and patient preferences for video or face-to-face consultations. A substantial number of adolescents who followed IMP-FITNET had one or more face-to-face consultations. A blended form of IMP-FITNET, in which different modalities of communication can be used, may have advantages and is in line with the current practice to combine internet interventions with face-to-face interaction with a therapist. One limitation is that during implementation of the FITNET treatment, technical options were expanded, for example, video consultations were integrated in the portal. The increasingly rapid development within software systems makes it more difficult to compare treatments designed at different time points. More research is necessary to inform when blended CBT is more effective than internet-based treatment alone. Further research also has to show whether blended care, with video consultations, is as cost-effective as FITNET with only email contact.

In conclusion, this study showed that IMP-FITNET is an effective and safe treatment for adolescents with CFS/ME in RCC. In RCC, the therapist can tailor the mode of delivery of the intervention to the needs of the individual patient.

Authors' Contributions

HK was the principal investigator of this study. EA and LNN were responsible for data gathering. EA, LNN, EEBVDS, and HK were responsible for the data analysis and for drafting the report. HK, SLN, and EMVDP designed and supervised the study, and revised the manuscript critically. All authors have read and approved the final submitted manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Posttreatment scores of the 226 patients in routine clinical care who fulfilled all the inclusion criteria of the Fatigue in Teenagers on the Internet randomized controlled trial.
[DOCX File , 14 KB - jmir_v23i8e24839_app1.docx]

References

1. Fukuda K, Straus SE, Hickie I, Sharpe MC, Dobbins JG, Komaroff A. The chronic fatigue syndrome: a comprehensive approach to its definition and study. International Chronic Fatigue Syndrome Study Group. Ann Intern Med 1994 Dec 15;121(12):953-959. [doi: 10.7326/0003-4819-121-12-199412150-00009] [Medline: 7978722]
2. Nijhof SL, Maijer K, Bleijenberg G, Uiterwaal CSPM, Kimpen JLL, van de Putte EM. Adolescent chronic fatigue syndrome: prevalence, incidence, and morbidity. Pediatrics 2011 May;127(5):e1169-e1175. [doi: 10.1542/peds.2010-1147] [Medline: 21502228]
3. Chalder T, Goodman R, Wessely S, Hotopf M, Meltzer H. Epidemiology of chronic fatigue syndrome and self reported myalgic encephalomyelitis in 5-15 year olds: cross sectional study. BMJ 2003 Sep 20;327(7416):654-655 [FREE Full text] [doi: 10.1136/bmj.327.7416.654] [Medline: 14500438]
4. Nijhof SL, Bleijenberg G, Uiterwaal CSPM, Kimpen JLL, van de Putte EM. Effectiveness of internet-based cognitive behavioural treatment for adolescents with chronic fatigue syndrome (FITNET): a randomised controlled trial. Lancet 2012 Apr 14;379(9824):1412-1418. [doi: 10.1016/S0140-6736(12)60025-7] [Medline: 22385683]
5. Nijhof LN, Nijhof SL, Bleijenberg G, Stellato RK, Kimpen JLL, Hulshoff Pol HE, et al. The impact of chronic fatigue syndrome on cognitive functioning in adolescents. Eur J Pediatr 2016 Feb 3;175(2):245-252 [FREE Full text] [doi: 10.1007/s00431-015-2626-1] [Medline: 26334394]
6. Lievesley K, Rimes KA, Chalder T. A review of the predisposing, precipitating and perpetuating factors in Chronic Fatigue Syndrome in children and adolescents. Clin Psychol Rev 2014 Apr;34(3):233-248. [doi: 10.1016/j.cpr.2014.02.002] [Medline: 24632047]
7. Stulemeijer M, de Jong LWAM, Fiselier TJW, Hoogveld SWB, Bleijenberg G. Cognitive behaviour therapy for adolescents with chronic fatigue syndrome: randomised controlled trial. BMJ 2005 Jan 01;330(7481):14 [FREE Full text] [doi: 10.1136/bmj.38301.587106.63] [Medline: 15585538]
8. Chalder T, Deary V, Husain K, Walwyn R. Family-focused cognitive behaviour therapy versus psycho-education for chronic fatigue syndrome in 11- to 18-year-olds: a randomized controlled treatment trial. Psychol Med 2010 Aug;40(8):1269-1279. [doi: 10.1017/S003329170999153X] [Medline: 19891804]
9. Baos S, Brigden A, Anderson E, Hollingworth W, Price S, Mills N, et al. Investigating the effectiveness and cost-effectiveness of FITNET-NHS (Fatigue In Teenagers on the interNET in the NHS) compared to Activity Management to treat paediatric chronic fatigue syndrome (CFS)/myalgic encephalomyelitis (ME): protocol for a randomised controlled trial. Trials 2018 Feb 22;19(1):136 [FREE Full text] [doi: 10.1186/s13063-018-2500-3] [Medline: 29471861]
10. Anderson E, Parslow R, Hollingworth W, Mills N, Beasant L, Gaunt D, et al. Recruiting Adolescents With Chronic Fatigue Syndrome/Myalgic Encephalomyelitis to Internet-Delivered Therapy: Internal Pilot Within a Randomized Controlled Trial. J Med Internet Res 2020 Aug 12;22(8):e17768 [FREE Full text] [doi: 10.2196/17768] [Medline: 32784188]
11. Rothwell P. Can overall results of clinical trials be applied to all patients? Lancet 1995 Jun 24;345(8965):1616-1619. [doi: 10.1016/s0140-6736(95)90120-5] [Medline: 7783541]
12. Garfield SL. Some problems associated with "validated" forms of psychotherapy. Clinical Psychology: Science and Practice 1996;3(3):218-229. [doi: 10.1111/j.1468-2850.1996.tb00073.x]
13. M.E. 2008: What progress? Solihull & South Birmingham ME Support Group. 2008. URL: https://ssb4mesupport.weebly.com/uploads/8/0/5/0/8050248/action_for_me_survey_2008.pdf [accessed 2021-07-28]
14. Heins MJ, Knoop H, Prins JB, Stulemeijer M, van der Meer JWM, Bleijenberg G. Possible detrimental effects of cognitive behaviour therapy for chronic fatigue syndrome. Psychother Psychosom 2010 Jun;79(4):249-256. [doi: 10.1159/000315130] [Medline: 20502065]
15. Dougall D, Johnson A, Goldsmith K, Sharpe M, Angus B, Chalder T, et al. Adverse events and deterioration reported by participants in the PACE trial of therapies for chronic fatigue syndrome. J Psychosom Res 2014 Jul;77(1):20-26 [FREE Full text] [doi: 10.1016/j.jpsychores.2014.04.002] [Medline: 24913337]
16. Reeves WC, Lloyd A, Vernon SD, Klimas N, Jason LA, Bleijenberg G, International Chronic Fatigue Syndrome Study Group. Identification of ambiguities in the 1994 chronic fatigue syndrome research case definition and recommendations for resolution. BMC Health Serv Res 2003 Dec 31;3(1):25 [FREE Full text] [doi: 10.1186/1472-6963-3-25] [Medline: 14702202]
17. Vercoulen JH, Swanink CM, Fennis JF, Galama JM, van der Meer JW, Bleijenberg G. Dimensional assessment of chronic fatigue syndrome. J Psychosom Res 1994 Jul;38(5):383-392. [doi: 10.1016/0022-3999(94)90099-x] [Medline: 7965927]
18. Lecrubier Y, Sheehan D, Weiller E, Amorim P, Bonora I, Sheehan KH, et al. The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: reliability and validity according to the CIDI. Eur. psychiatr 2020 Apr 16;12(5):224-231. [doi: 10.1016/s0924-9338(97)83296-8]
19. Knoop H, Prins JB, Moss-Morris R, Bleijenberg G. The central role of cognitive processes in the perpetuation of chronic fatigue syndrome. J Psychosom Res 2010 May;68(5):489-494. [doi: 10.1016/j.jpsychores.2010.01.022] [Medline: 20403509]

20. Nijhof SL, Bleijenberg G, Uiterwaal CS, Kimpen JL, van de Putte EM. Fatigue In Teenagers on the interNET--the FITNET Trial. A randomized clinical trial of web-based cognitive behavioural therapy for adolescents with chronic fatigue syndrome: study protocol. [ISRCTN59878666]. BMC Neurol 2011 Feb 19;11(1):23 [FREE Full text] [doi: 10.1186/1471-2377-11-23] [Medline: 21333021]
21. Raat H, Landgraf JM, Bonsel GJ, Gemke RJBJ, Essink-Bot ML. Reliability and validity of the child health questionnaire-child form (CHQ-CF87) in a Dutch adolescent population. Qual Life Res 2002 Sep;11(6):575-581. [doi: 10.1023/a:1016393311799] [Medline: 12206578]
22. Collin SM, Nuevo R, van de Putte EM, Nijhof SL, Crawley E. Chronic fatigue syndrome (CFS) or myalgic encephalomyelitis (ME) is different in children compared to in adults: a study of UK and Dutch clinical cohorts. BMJ Open 2015 Oct 28;5(10):e008830 [FREE Full text] [doi: 10.1136/bmjopen-2015-008830] [Medline: 26510728]
23. VanderZee KI, Sanderman R, Heyink JW, de Haes H. Psychometric qualities of the RAND 36-Item Health Survey 1.0: a multidimensional measure of general health status. Int J Behav Med 1996;3(2):104-122. [doi: 10.1207/s15327558ijbm0302_2] [Medline: 16250758]
24. Chapman AL, Hadfield M, Chapman CJ. Qualitative research in healthcare: an introduction to grounded theory using thematic analysis. J R Coll Physicians Edinb 2015;45(3):201-205. [doi: 10.4997/JRCPE.2015.305] [Medline: 26517098]
25. Pedersen AB, Mikkelsen EM, Cronin-Fenton D, Kristensen NR, Pham TM, Pedersen L, et al. Missing data and multiple imputation in clinical epidemiological research. Clin Epidemiol 2017;9:157-166 [FREE Full text] [doi: 10.2147/CLEP.S129785] [Medline: 28352203]
26. Jacobson N, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991 Feb;59(1):12-19. [doi: 10.1037//0022-006x.59.1.12] [Medline: 2002127]
27. Tummers M, Knoop H, van Dam A, Bleijenberg G. Implementing a minimal intervention for chronic fatigue syndrome in a mental health centre: a randomized controlled trial. Psychol Med 2012 Oct;42(10):2205-2215. [doi: 10.1017/S0033291712000232] [Medline: 22354999]
28. Worm-Smeitink M, Janse A, van Dam A, Evers A, van der Vaart R, Wensing M, et al. Internet-Based Cognitive Behavioral Therapy in Stepped Care for Chronic Fatigue Syndrome: Randomized Noninferiority Trial. J Med Internet Res 2019 Mar 14;21(3):e11276 [FREE Full text] [doi: 10.2196/11276] [Medline: 30869642]
29. Kyriacou DN, Lewis RJ. Confounding by Indication in Clinical Research. JAMA 2016 Nov 01;316(17):1818-1819. [doi: 10.1001/jama.2016.16435] [Medline: 27802529]

Abbreviations

CBT: cognitive behavior therapy
CFS: chronic fatigue syndrome
CHQ-87: Child Health Questionnaire-87
CIS-20: Checklist Individual Strength-20
FITNET: Fatigue in Teenagers on the Internet
I-CBT: internet-based cognitive behavior therapy
IMP-FITNET: implemented Fatigue in Teenagers on the Internet
ME: myalgic encephalomyelitis
RCC: routine clinical care
RCT: randomized controlled trial

Edited by R Kukafka; submitted 08.10.20; peer-reviewed by L Larun, A Lundervold, B Johansson; comments to author 09.02.21; revised version received 23.02.21; accepted 30.06.21; published 13.08.21.

Please cite as:
Albers E, Nijhof LN, Berkelbach van der Sprenkel EE, van de Putte EM, Nijhof SL, Knoop H
Effectiveness of Internet-Based Cognitive Behavior Therapy (Fatigue in Teenagers on the Internet) for Adolescents With Chronic Fatigue Syndrome in Routine Clinical Care: Observational Study
J Med Internet Res 2021;23(8):e24839
URL: https://www.jmir.org/2021/8/e24839
doi:10.2196/24839
PMID:34397389

©Eline Albers, Linde N Nijhof, Emma E Berkelbach van der Sprenkel, Elise M van de Putte, Sanne L Nijhof, Hans Knoop. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

New Internet-Based Warfarin Anticoagulation Management Approach After Mechanical Heart Valve Replacement: Prospective, Multicenter, Randomized Controlled Trial

Zhihui Zhu[1,2], MM; Chenyu Li[3], MM; Jinglun Shen[1], BA; Kaisheng Wu[1], BA; Yuehuan Li[1], MD; Kun Liu[1], MM; Fan Zhang[1], MM; Zhenhua Zhang[1,4], MM; Yan Li[1], MD; Jie Han[1], BA; Ying Qin[1], MA; Yu Yang[5], MA; Guangpu Fan[6], PhD; Huajun Zhang[7], DPhil, MD; Zheng Ding[8], MA; Dong Xu[5], MD; Yu Chen[6], MD; Yingli Zheng[8], DPharm; Zhe Zheng[9], PhD, MD; Xu Meng[1], MD; Haibo Zhang[1], PhD, MD

[1]Department of Cardiac Surgery, Beijing Anzhen Hospital, Capital Medical University, Beijing, China

[2]Ludwig Maximilian University of Munich, Munich, Germany

[3]Renal Division, Department of Medicine IV, Ludwig Maximilian University of Munich, Munich, Germany

[4]Department of Cardiovascular Surgery, Beijing Luhe Hospital,, Capital Medical University, Beijing, China

[5]Department of Cardiovascular Surgery, Beijing Xuanwu Hospital, Beijing, China

[6]Department of Cardiovascular Surgery, Peking University People’s Hospital, Beijing, China

[7]Department of Cardiovascular Surgery, PLA General Hospital, Beijing, China

[8]Department of Pharmacy, Fuwai Hospital, National Center for Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

[9]Department of Cardiovascular Surgery, Fuwai Hospital, National Center for Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

Corresponding Author:
Haibo Zhang, PhD, MD
Department of Cardiac Surgery
Beijing Anzhen Hospital
Capital Medical University
2 Anzhen Rd
Beijing, 100029
China
Phone: 86 1 337 010 3561
Email: zhanghb2318@163.com

Abstract

Background: Mechanical heart valve replacement (MHVR) is an effective method for the treatment of severe heart valve disease; however, it subjects patient to lifelong warfarin therapy after MHVR with the attendant risk of bleeding and thrombosis. Whether internet-based warfarin management reduces complications and improves patient quality of life remains unknown.

Objective: This study aimed to compare the effects of internet-based warfarin management and the conventional approach in patients who received MHVR in order to provide evidence regarding alternative strategies for long-term anticoagulation.

Methods: This was a prospective, multicenter, randomized, open-label, controlled clinical trial with a 1-year follow-up. Patients who needed long-term warfarin anticoagulation after MHVR were enrolled and then randomly divided into conventional and internet-based management groups. The percentage of time in the therapeutic range (TTR) was used as the primary outcome, while bleeding, thrombosis, and other events were the secondary outcomes.

Results: A total of 721 patients were enrolled. The baseline characteristics did not reach statistical differences between the 2 groups, suggesting the random assignment was successful. As a result, the internet-based group showed a significantly higher TTR (mean 0.53, SD 0.24 vs mean 0.46, SD 0.21; *P*<.001) and fraction of time in the therapeutic range (mean 0.48, SD 0.22 vs mean 0.42, SD 0.19; *P*<.001) than did those in the conventional group. Furthermore, as expected, the anticoagulation complications, including the bleeding and embolic events had a lower frequency in the internet-based group than in the conventional group (6.94% vs 12.74%; *P*=.01). Logistic regression showed that internet-based management increased the TTR by 7% (odds ratio

[OR] 1.07, 95% CI 1.05-1.09; $P<.001$) and reduced the bleeding and embolic risk by 6% (OR 0.94, 95% CI 0.92-0.96; $P=.01$). Moreover, low TTR was found to be a risk factor for bleeding and embolic events (OR 0.87, 95% CI 0.83-0.91; $P=.005$).

Conclusions: The internet-based warfarin management is superior to the conventional method, as it can reduce the anticoagulation complications in patients who receive long-term warfarin anticoagulation after MHVR.

Trial Registration: Chinese Clinical Trial Registry ChiCTR1800016204; http://www.chictr.org.cn/showproj.aspx?proj=27518

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2019-032949

(J Med Internet Res 2021;23(8):e29529) doi:10.2196/29529

KEYWORDS

RCT; warfarin; telemedicine; TTR; complication

Introduction

Heart valve replacement is recommended for patients with severe heart valve disease and is performed in several hundred thousand patients worldwide each year [1]. As a replacement valve for humans, a mechanical valve is more durable than a bioprosthetic valve but requires lifelong anticoagulation therapy [2]. Over 98% of young Chinese patients who need heart valve replacement receive mechanical valves [3] but face the risk of thrombosis on the valves and ensuing embolism [4]; thus, lifelong anticoagulation is required and recommended by guidelines [5]. Warfarin, as an anticoagulation drug, is widely used in the prevention of various thromboembolic events [6], but it is not easy to control due to the narrow therapeutic range and patients' heterogeneity. Therefore, the dosage of warfarin needs to be adjusted accordingly by the international normalized ratio (INR) [7,8]. Time in therapeutic range (TTR), as the percentage of time the patient's INR is within the target range, has been used in clinical research to measure the adequacy of warfarin therapy and shows a significant correlation with anticoagulation outcomes [9,10]. Well-managed warfarin anticoagulation is effective in reducing complications, such as bleeding and thrombosis, in patients with a mechanical heart valve [11]. However, in clinical practice, patients' anticoagulation management is not conducted in ideal fashion.

The conventional way to manage anticoagulation is face to face [12], but the quality of this approach depends highly on the patients' compliance, resulting in difficulty in management and increased risk of adverse events [13,14]. Chinese cardiac surgeons tend to focus more on inpatient treatment, but follow-up is often ignored, especially in central and western China, where there is a lack of professional cardiac surgery–related health care [15]. Furthermore, the extensive follow-up required for warfarin management adds to the workload of specialist hospitals [16,17]. Therefore, current warfarin management needs to be urgently improved.

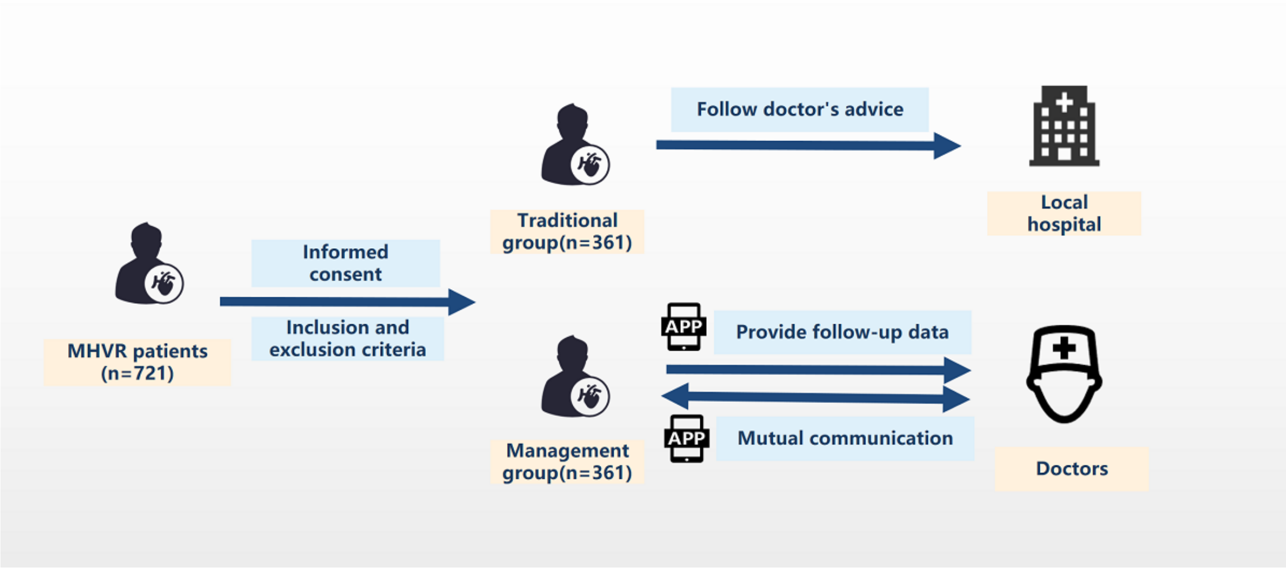
The internet provides fast and convenient communication, with integration into health care delivery systems presenting exciting opportunities for improving care promotion, disease prevention, and value-based clinical medicine [18]. Previous studies have reported several online warfarin management systems and have demonstrated their benefits for patients after heart valve surgery [19-21]. However, the capacity of these telemedicine apps to improve the quality of anticoagulation remains unclear, and carefully designed prospective trials are needed to provide reliable evidence. Therefore, in this large scale, prospective, multicenter, randomized controlled trial, we aimed to study if internet-based warfarin management increases TTR and reduces the risk of anticoagulation complications in patients after mechanical heart valve replacement. Our study is the first randomized controlled trial to explore the effectiveness of internet-based warfarin anticoagulation management in patients after mechanical heart valve replacement.

Methods

Study Design

The clinical trial design has been described previously [16]. In brief, the study was a prospective, multicenter, randomized, open-label, controlled trial. The patient flow chart is shown in Figure 1. The project was organized and implemented by Beijing Anshan Hospital affiliated with Capital Medical University. Five top cardiac centers in China participated in the clinical trial. The Peking University Clinical Research Institute has created a data committee to evaluate the data quality and supervise data collecting.

Figure 1. Patient flow chart. MHVR mechanical heart valve replacement.



All participants provided written informed consent. The study protocol was approved by the Ethics Committee of Beijing Anshan Hospital, Capital Medical University, and all other centers accepted the central ethics approval or obtained local approval by internal ethics committees.

Participants

Inclusion criteria were as follows: patients had a normal psychological state and were aged 18 to 65 years, patients or guardians agreed to the study plan and signed the informed consent form, long-term anticoagulation was required after mechanical heart valve replacement, and patients and their families could effectively operate smartphones. The exclusion criteria were the following: patients with emergency surgery; patients treated with cardiac bypass surgery; patients with severe chronic diseases, such as malignant tumors, hepatic cirrhosis, diabetes, cerebral hemorrhage (cerebral infarction), convalescence, thyroid dysfunction, and respiratory failure; patients with severe renal insufficiency (endogenous creatinine clearance rate ≤20 mL/min); patients with severe hepatic insufficiency (Child-Pugh ≥10); patients with severe heart failure (cardiac function New York Heart Association grade IV); patients with postoperative infective endocarditis; patients with severe postoperative complications and poor prognosis; and patients who were unable to comply with this study.

Intervention

In our study, the conventional management was conducted by laboratory testing and drug dose adjustment at the hospital as instructed by a doctor, while internet-based management was performed via a mobile user interface medical network follow-up platform. All patients in both groups received training and education on warfarin anticoagulation before discharge, which included information on drug interactions, diet, compliance with dosing, and the importance of INR testing.

In the conventional anticoagulation management model, patients underwent INR tests at the local hospital. After reviewing the laboratory results, the outpatient cardiologist evaluated the INR results and decided whether to adjust the warfarin dosage until the next patient visit. Patients were asked to record the results of every examination and note the physical symptoms related to warfarin anticoagulation, such as bleeding and embolism, to prepare for follow-up discharge.

In the early stages of the study, we developed an internet-based follow-up system for clinical use after cardiac surgery [22]. In the new warfarin management group, online registration information was generated by researchers and sent to patients, who completed the rest of the follow-up registration. The app software for the internet-based follow-up management system has been described previously [16]. All patients and their families attended at least 3 follow-up training sessions, including watching an educational video and installation and simulation of the follow-up software and mobile apps. Patients were asked to upload the INR and other laboratory test results to the apps.

Randomization, Blinding, and Grouping

The central stratified randomization method was used in this study. Patients who met the inclusion and not exclusion criteria and consented to participate in this study were randomized in a 1:1 ratio into the conventional and internet-based management groups. Patients were aware of their grouping, whereas the researchers conducting the end point evaluations were blinded to the groupings.

Anticoagulation Treatment Strategies and Follow-Up

The INR was measured every 3 to 5 days for 1 month after discharge. After the patient’s INR results stabilized, the INR was measured every 10 to 14 days from 1 to 3 months after discharge, once a month from 3 to 6 months, and once every 2 months from 6 to 12 months after discharge. The target INR window ranged from 1.8 to 2.5 (1.8-2.2 for aortic valve replacement, 2.0-2.5 for mitral valve replacement, and 2.0-2.5 for double valve replacement).

Patients were followed up at 3 months (30 days before and after), 6 months (30 days before and after), 9 months (30 days before and after), and 12 months (30 days before and after) after discharge.

At each follow-up point, patients in the traditional group were asked to bring their original hospitalization records and all medical examination results after discharge, and the researchers completed the case report form. Patients were followed up by telephone if they could not return to the hospital at the scheduled time points.

For patients in the new warfarin group, the researchers checked the data results uploaded by the patients in the new follow-up system app and completed the case report form at the follow-up time points.

End Point

The primary end point of this study was the TTR. The TTR was measured by Roosendaal linear interpolation [10]. The secondary end point was the incidence of anticoagulation-related embolism and bleeding. Embolism includes any embolic event that occurs in the absence of infection after the immediate perioperative period. Embolism may manifest as a neurologic event or a noncerebral embolic event. Severe bleeding associated with anticoagulation includes any episode of major internal or external bleeding that causes death, hospitalization, or permanent injury or that necessitates transfusion. General bleeding related to anticoagulation refers to bleeding of the nasal cavity and gums, skin ecchymosis, menorrhea, hematuria, melena, and so on.

Statistical Analyses

Continuous variables are expressed as the mean and SD or median and IQR; meanwhile, categorical variables are expressed as frequencies and percentages. Baseline characteristics were compared between the 2 groups by the chi-square, *t*, and Mann-Whitney tests as appropriate. The Mann-Whitney test was used to determine the difference in the TTR between the conventional anticoagulation management model group and the internet-based anticoagulation model group. The incidence of adverse events (bleeding or thrombosis) was compared by the chi-square test. A 2-sided *P* value <0.05 was regarded as statistically significant. All analyses were performed using R version 3.4.2 (The R Foundation).

Results

Characteristics of the Patients

A total of 721 patients from 5 top Chinese cardiovascular centers were enrolled between July 2018 and September 2019 and then randomly assigned to a conventional group (n=361) or an internet-based warfarin management group (n=360). The average ages of the conventional and internet-based group were 49.59 (SD 9.46) and 50.6 (SD 9.65), respectively (*P*=0.16); the percentage of female patients in the 2 groups was 38.9% (140/360) and 39.1% (141/361), respectively (*P*=0.96). The other baseline characteristics also did not reach statistical differences between the 2 groups (Table 1), suggesting the patients' random assignment was successful.

Table 1. Patient characteristics.

Characteristics	Internet-based group (n=360)	Conventional group (n=361)	*P* value
Demographics			
Age (years), mean (SD[a])	49.59 (9.46)	50.6 (9.65)	.16
Female sex, n (%)	140 (38.89)	141 (39.06)	.96
Risk factors			
Hypertension, n (%)	92 (25.56)	99 (27.42)	.57
Atrial fibrillation, n (%)	69 (19.17)	71 (19.67)	.87
Angina, n (%)	11 (3.06)	13 (3.6)	.68
Smoking status			.55
Smokers, n (%)	67 (18.61)	63 (17.45)	
Nonsmokers, n (%)	224 (62.22)	233 (64.54)	
Former smokers, n (%)	69 (19.17)	65 (18.01)	
Clinical data, mean (SD)			
BMI[b], kg/m^2	24.23 (3.34)	24.52 (3.52)	.25
Preoperative resting heart rate (times/minute)	78.46 (15.41)	78.73 (15.17)	.81
Preoperative EF[c] (%)	61.18 (7.03)	60.88 (7.59)	.58
Preoperative platelet (10^9/L)	201.57 (55.97)	201.35 (56.45)	.96
Preoperative ALT[d] (U/L)	25.21 (23.71)	25.47 (20.56)	.88
Preoperative AST[e] (U/L)	27.3 (22.28)	27.49 (16.34)	.90
Preoperative BUN[f] (mmol/L)	6.51 (2.29)	6.54 (2.23)	.86
Preoperative Cr[g] (umol/L)	79.73(19.76)	79.39 (20.77)	.82
Postoperative resting heart rate (times/minute)	87.51(15.81)	87.02 (15.89)	.67
Postoperative ALT (U/L)	44.77 (46.03)	40.9 (36.5)	.21
Postoperative AST (U/L)	38.57 (26)	35.84 (22.96)	.14
Postoperative BUN (mmol/L)	7.85 (3.27)	7.88 (3.33)	.89
Postoperative Cr (umol/L)	74.37 (25.72)	76.19 (23.61)	.32

[a] SD: standard deviation.

[b] BMI: body mass index.

[c] EF: ejection fraction.

[d] ALT: alanine transaminase.

[e] AST: aspartate transaminase.

[f] BUN: blood urea nitrogen.

[g] Cr: creatinine.

Internet-Based Warfarin Management Increased TTR

In order to study the effect of internet-based management on warfarin anticoagulation, we used TTR as the primary end point. The TTR was significantly higher in the internet-based group than in the conventional group (internet: mean 0.53, SD 0.24; conventional: mean 0.46, SD 0.21; *P*<.001). Of the 360 patients in the internet-based group, the number of patients with a TTR in the range of 0%-30%, 30%-60%, and 60%-100% was 61 (16.94%), 156 (43.33%), and 143 (39.72%), respectively; meanwhile, of the 361 patients in the conventional group, the number was 77 (21.33%), 189 (52.35%), and 95 (26.32%), respectively. As expected, the internet-based group yielded a higher fraction of TTR than did the traditional group (internet: mean 0.48, SD 0.22; conventional: mean 0.42, SD 0.19; *P*<.001). Logistic regression indicated that the internet-based management increased the TTR by 7% (odds ratio [OR] 1.07, 95% CI 1.05-1.09; *P*<.001; Table 2). Our results suggest that internet-based warfarin management is better than the conventional method according to increased TTR, while the mean INR values (internet: mean 2.13, SD 0.87; conventional: mean 2.14, SD 1.10; *P*=0.87) were not different between the 2 groups (Table 3).

Table 2. Logistic regression of time in the therapeutic range.

Factor	β	OR[b] (95% CI)	*t*	*P* value
Age	0 (0 to 0)	1 (1 to 1)	1.35	.18
Female sex	–0.01 (–0.03 to 0.01)	0.99 (0.98 to 1.01)	–0.44	.66
Hypertension	0.05 (0.03 to 0.07)	1.05 (1.03 to 1.07)	2.74	.01
Atrial fibrillation	–0.06 (–0.08 to –0.03)	0.95 (0.93 to 0.97)	–2.56	.01
Angina	0.06 (0.01 to 0.11)	1.06 (1.01 to 1.11)	1.25	.21
Smoking status	0.01 (0 to 0.02)	1.01 (1 to 1.02)	1.06	.29
BMI[c]	0.01 (0 to 0.01)	1.01 (1 to 1.01)	2.71	.01
Preoperative resting heart rate	0 (0 to 0)	1 (1 to 1)	–0.22	.83
Preoperative EF[d]	0 (0 to 0)	1 (1 to 1)	–0.38	.70
Preoperative platelet	0 (0 to 0)	1 (1 to 1)	–0.37	.71
Preoperative ALT[e]	0 (0 to 0)	1 (1 to 1)	–0.73	.46
Preoperative AST[f]	0 (0 to 0)	1 (1 to 1)	–0.85	.39
Preoperative BUN[g]	0 (0 to 0)	1 (1 to 1)	–0.21	.83
Preoperative Cr[h]	0 (0 to 0)	1 (1 to 1)	–2.29	.02
Postoperative resting heart rate	0 (0 to 0)	1 (1 to 1)	–0.29	.77
Postoperative ALT	0 (0 to 0)	1 (1 to 1)	1.48	.14
Postoperative AST	0 (0 to 0)	1 (1 to 1)	1.46	.14
Postoperative BUN	0.01 (0.01 to 0.01)	1.01 (1.01 to 1.01)	3.02	.002
Postoperative Cr	0 (0 to 0)	1 (1 to 1)	–1.29	.20
Group	0.07 (0.05 to 0.08)	1.07 (1.05 to 1.09)	3.96	<.001
FTTR[i]	0.97 (0.95 to 0.99)	2.63 (2.57 to 2.69)	4.46	<.001
General bleeding event	–0.08 (–0.11 to –0.05)	0.92 (0.89 to 0.95)	–2.78	.01
Severe bleeding event	–0.11 (–0.21 to –0.02)	0.89 (0.81 to 0.98)	–1.19	.23
All bleeding events	–0.09 (–0.12 to –0.06)	0.92 (0.89 to 0.94)	–3.05	.002
Neurologic embolic event	0.08(–0.08 to 0.24)	1.08 (0.92 to 1.27)	0.49	.62
Noncerebral embolic event	0.09 (–0.14 to 0.32)	1.09 (0.87 to 1.38)	0.39	.70
All embolic events	0.08 (–0.05 to 0.22)	1.09 (0.95 to 1.24)	0.62	.53
All bleeding and embolic events	–0.08 (–0.11 to –0.05)	0.92 (0.9 to 0.95)	–2.85	.005
Revisit to hospital due to secondary end point	–0.05 (–0.12 to 0.02)	0.95 (0.88 to 1.02)	–0.69	.49
Death due to secondary end point	0.03 (–0.14 to 0.19)	1.03 (0.87 to 1.21)	0.16	.87

[a] TTR: time in the therapeutic range.

[b] OR: odds ratio.

[c] BMI: body mass index.

[d] EF: ejection fraction.

[e] ALT: alanine transaminase.

[f] AST: aspartate transaminase.

[g] BUN: blood urea nitrogen.

[h] Cr: creatinine.

[i] FTTR: fraction of time in therapeutic range.

Table 3. Primary end point.

End point	Internet-based group (n=360)	Traditional group (n=361)	*P* value
INR[a], mean (SD[b])	2.13 (0.87)	2.14 (1.10)	.87
TTR[c], mean (SD)	0.53 (0.24)	0.46 (0.21)	<.001
0%-30%, n (%)	61 (16.94)	77 (21.33)	N/A[d]
30%-60%, n (%)	156 (43.33)	189 (52.35)	N/A
60%-100%, n (%)	143 (39.72)	95 (26.32)	N/A
FTTR[e], mean (SD)	0.48 (0.22)	0.42 (0.19)	<.001

[a] INR: international normalized ratio.

[b] SD: standard deviation.

[c] TTR: time in the therapeutic range.

[d] N/A: not applicable.

[e] FTTR: fraction of time in therapeutic range.

Internet-Based Warfarin Management Reduced the Risk of Adverse Events

A lower risk of stroke and bleeding can reached by maximizing the TTR. We further used the bleeding and embolic events as a secondary end point to investigate the difference between the internet-based and conventional groups. The incidence of all bleeding and embolic events (6.94% vs 12.74%; *P*=.009) was lower in the internet-based group than in the conventional group (Table 4). Logistic regression showed that internet-based management reduced the bleeding and embolic risk by 6% (OR 0.94, 95% CI 0.92-0.96; *P*=.01). Moreover, low TTR was a risk factor for bleeding and embolic events (OR 0.87, 95% CI 0.83-0.91; *P*=.005; Table 5). The results showed that internet-based warfarin management not only increased TTR but also reduced the risk of adverse events of warfarin anticoagulation.

Table 4. Adverse events.

Adverse event	Internet-based group (n=360)	Traditional group (n=361)	*P* value
General bleeding event, n (%)	22(6.11)	40(11.08)	.02
Severe bleeding event, n (%)	2(0.56)	4(1.11)	.41
All bleeding events, n (%)	24(6.67)	44(12.19)	.01
Neurologic embolic event, n (%)	1(0.28)	1(0.28)	>.99
Noncerebral embolic event, n (%)	0(0)	1(0.28)	.32
All embolic events, n (%)	1(0.28)	2(0.55)	.56
All bleeding and embolic events, n (%)	25(6.94)	46(12.74)	.01
Revisit to hospital due to secondary end point, n (%)	4(1.11)	6(1.66)	.53
Death due to secondary end point, n (%)	0(0)	2(0.55)	.87

Table 5. Logistic regression of all bleeding and embolic events.

Characteristic	β	OR[a] (95% CI)	*t*	*P* value
Age	0 (0 to 0)	1 (1 to 1)	0.51	.61
Female sex	–0.02 (–0.04 to 0)	0.98 (0.96 to 1)	–0.94	.35
Hypertension	0.02 (0 to 0.05)	1.02 (1 to 1.05)	0.90	.37
Atrial fibrillation	0.04 (0.01 to 0.07)	1.04 (1.01 to 1.07)	1.33	.18
Angina	0.11 (0.05 to 0.18)	1.12 (1.05 to 1.19)	1.84	.07
Smoking status	0.03 (0.01 to 0.04)	1.03 (1.01 to 1.04)	2.04	.04
BMI[b]	0 (0 to 0)	1 (1 to 1)	–0.53	.60
Preoperative resting heart rate	0 (0 to 0)	1 (1 to 1)	–0.31	.75
Preoperative EF[c]	0 (0 to 0)	1 (1 to 1)	–0.99	.32
Preoperative platelet	0 (0 to 0)	1 (1 to 1)	–1.18	.24
Preoperative ALT[d]	0 (0 to 0)	1 (1 to 1)	0.03	.98
Preoperative AST[e]	0 (0 to 0)	1 (1 to 1)	0.67	.51
Preoperative BUN[f]	0.01 (0 to 0.01)	1.01 (1 to 1.01)	1.81	.07
Preoperative Cr[g]	0 (0 to 0)	1 (1 to 1)	1.70	.09
Postoperative resting heart rate	0 (0 to 0)	1 (1 to 1)	–1.06	.29
Postoperative ALT	0 (0 to 0)	1 (1 to 1)	–0.55	.58
Postoperative AST	0 (0 to 0)	1 (1 to 1)	–0.55	.58
Postoperative BUN	0 (–0.01 to 0)	1 (0.99 to 1)	–1.09	.28
Postoperative Cr	0 (0 to 0)	1 (1 to 1)	1.18	.24
Group	–0.06 (–0.08 to –0.04)	0.94 (0.92 to 0.96)	–2.62	.01
TTR[h]	–0.14 (–0.19 to –0.09)	0.87 (0.83 to 0.91)	–2.85	.005
FTTR[i]	–0.2 (–0.25 to –0.14)	0.82 (0.78 to 0.87)	–3.66	<.001
Revisit to hospital due to secondary end point	0.91 (0.83 to 1)	2.49 (2.28 to 2.73)	10.31	<.001
Death due to secondary end point	0.9 (0.7 to 1.11)	2.47 (2 to 3.04)	4.33	<.001

[a] OR: odds ratio.

[b] BMI: body mass index.

[c] EF: ejection fraction.

[d] ALT: alanine transaminase.

[e] AST: aspartate transaminase.

[f] BUN: blood urea nitrogen.

[g] Cr: creatinine.

[h] TTR: time in the therapeutic range.

[i] FTTR: fraction of time in therapeutic range.

Discussion

In this study, there was a significantly higher level of TTR in the population that was administered with the internet-based anticoagulation management than in the population treated with conventional warfarin management (0.53 vs 0.46; *P*<.001) in a real-world medical setting. Management of warfarin anticoagulation by a specialized, telemedicine–based service was able to substantially improve the quality of anticoagulation therapy. Our research findings showed that the new telemedical warfarin management had lower incidences of general bleeding events (22/260, 6.11% vs 40/361, 11.08%; *P*=.02), all bleeding events (24/360, 6.67% vs 44/361, 12.19%; *P*=.01), and all bleeding and embolic events than did the conventional group. In addition, single factor logistic regression of TTR and all bleeding and embolic events demonstrated the superiority of the new internet-based warfarin management as well as the relationship between TTR and anticoagulation complications in the 2 groups.

Patients who receive mechanical heart valve replacement need lifelong anticoagulation therapy, and their INR is conventionally measured to adjust the anticoagulation strength and the dose of anticoagulation drugs. However, adjusting the warfarin dosage is a challenging task since there is heterogeneity in patients' warfarin dosing [23]. It has been reported that Chinese patients require a lower dose of warfarin than do White patients although the intensity of the anticoagulation is comparable [24,25]. As Chinese patients have a lower warfarin requirement, there are concerns about whether the target INR range for Western populations also provides optional anticoagulation in Chinese patients. In one study, You et al concluded that an INR of 1.8 to 2.4 appeared to be associated with the lowest incidence of major bleeding or thromboembolic events in a cohort of Chinese patients receiving warfarin therapy for moderate-intensity anticoagulation [26]. Unfortunately, there is no guideline for the target value of INR in China. At the seminar before the start of the project, after rigorous discussion, the researchers from all subcenters finally determined the target INR window of 1.8 to 2.5 (1.8-2.2 for aortic valve replacement, 2.0-2.5 for mitral valve replacement, and 2.0-2.5 for a double valve replacement) according to the actual clinical situation and previous studies in our centers [23,27-30]. The effectiveness of warfarin anticoagulation therapy is usually expressed as TTR. Due to the relatively narrow INR window, the TTR in the 2 groups was correspondingly lower than that reported in other warfarin management studies [19,20,31,32].

The finding of high levels of TTR and low incidences of anticoagulation complications in the internet-based management group might be, at least in part, attributable to the education, reminders, and convenient doctor-patient communication of the new management platform, which has been described previously [16,22]. Led by the director of surgery and including surgeons, trained physician assistants, staff nurses, and pharmacists, the new anticoagulation management model was developed and combined health science promotion and education, portable coagulation indicator monitoring, warfarin-related gene monitoring, warfarin dosage predication, access to a professional outpatient clinic, and a publicity manual. Unlike other studies in which only the patients use the app, our mobile apps are specific to doctors and patients. The app for doctors was distributed to cardiovascular surgeons from the collaborating units participating in the study by the lead unit of this study; the app for patients was distributed to patients enrolled in this study by doctors participating in this study. The doctor's app contained a patient management module and a professional learning module that allowed timely patient contact. The patient's app contained the patients' personal information, including hospitalization information and discharge matters requiring attention. The patient's app was mainly divided into a health management module, doctor consultation module, and medical science popularization module. The app regularly reminded the patients of the time of the INR test. If patients missed the test, the software would warn them. This new platform completed the collection of clinical data and provided reliable and quality follow-up services and warfarin management for patients undergoing lifelong warfarin coagulation.

Telemedicine, a term used interchangeably with telehealth, is the distribution of health-related services and information via electronic information and telecommunication technologies [33]. Over the past four decades, telemedicine has become an increasingly effective alternative to traditional medicine and has evolved into an integrated technology used in hospitals and clinics [34]. As the public becomes more adept at using the internet and smartphone in all aspects of daily life, evolving app in health care will change the way in which patients and doctors interact [35]. A few previous studies explored several internet-based warfarin management approaches and showed the advantages of the new anticoagulation model [32,36-38]. However, their conclusions were not completely consistent.

Based on the modern communication technologies, a new type of anticoagulation mode for cardiac surgery was established via the "Internet+Medical" management mode, supported by knowledge-based clinical medical diagnosis and treatment. This prospective, multicenter, randomized, open-label, controlled trial fills a deficit in research and adds to the solid evidence which exists regarding optimal strategies for patients with an extended period of warfarin anticoagulation. The new internet-based anticoagulation management model has the potential to provide considerable social and economic benefits in terms of further improving the long-term survival and quality of life of patients undergoing lifelong anticoagulation therapy.

There were also some limitations to this study. First, for the anticoagulation complications, the sample size was relevantly small, and the results may not be representative of all populations with anticoagulation therapy; thus, more prospective, large sample, randomized studies are needed to confirm our findings. Second, patients' age, education level, and disease severity might have affected their compliance, which also might have influenced the quality of anticoagulation after surgery. Third, we did not compare the cost in the 2 management groups or the difference between urban and rural patients. In future studies and follow-up research, we will compare the associated medical costs during follow-up and rehabilitation under the 2 warfarin management models.

Internet-based warfarin management is superior to conventional management, as it can reduce the anticoagulation complications in patients who receive long-term warfarin anticoagulation after mechanical heart valve replacement.

Acknowledgments

The work is supported by the Beijing Science and Technology Commission of China (New warfarin anticoagulation management model after heart valve surgery; #D171100002917003). The authors are thankful for all the researchers from the 4 participating centers, Peking University Clinical Research Institute, and the technical support of Digital Health China Technologies Co, Ltd.

Authors' Contributions

HZ, ZZ, and YL conceived the study idea. CL, JS, KW, KL, FZ, ZZ, YL, JH, YQ, YY, GF, HZ, ZD, DX, YC, RW, YZ, ZZ, and XM made substantial contributions to the development of the study. ZZ and HZ drafted the manuscript, and all the authors contributed to the critical revisions of the paper. The final manuscript was read and approved by all the authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2005 KB - jmir_v23i8e29529_app1.pdf]

References

1. Eikelboom JW, Connolly SJ, Brueckmann M, Granger CB, Kappetein AP, Mack MJ, RE-ALIGN Investigators. Dabigatran versus warfarin in patients with mechanical heart valves. N Engl J Med 2013 Sep 26;369(13):1206-1214. [doi: 10.1056/NEJMoa1300615] [Medline: 23991661]
2. Stassano P, Di Tommaso L, Monaco M, Iorio F, Pepino P, Spampinato N, et al. Aortic valve replacement: a prospective randomized evaluation of mechanical versus biological valves in patients ages 55 to 70 years. J Am Coll Cardiol 2009 Nov 10;54(20):1862-1868 [FREE Full text] [doi: 10.1016/j.jacc.2009.07.032] [Medline: 19892237]
3. Chen Q, Dong L, Dong Y, Zhao S, Fu B, Wang Y, et al. Security and cost comparison of INR self-testing and conventional hospital INR testing in patients with mechanical heart valve replacement. J Cardiothorac Surg 2015 Jan 16;10:4 [FREE Full text] [doi: 10.1186/s13019-015-0205-1] [Medline: 25592732]
4. Nishimura R, Otto CM, Bonow RO, Carabello BA, Erwin JP, Fleisher LA, et al. 2017 AHA/ACC focused update of the 2014 AHA/ACC Guideline for the Management of Patients With Valvular Heart Disease: a report of the American College of Cardiology/American Heart Association Task Force on clinical practice guidelines. J Am Coll Cardiol 2017 Jul 11;70(2):252-289 [FREE Full text] [doi: 10.1016/j.jacc.2017.03.011] [Medline: 28315732]
5. Kearon C, Akl EA, Comerota AJ, Prandoni P, Bounameaux H, Goldhaber SZ, et al. Antithrombotic therapy for VTE disease: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. Chest 2012 Feb;141(2 Suppl):e419S-e496S [FREE Full text] [doi: 10.1378/chest.11-2301] [Medline: 22315268]
6. Qiu S, Wang N, Zhang C, Gu Z, Qian Y. Anticoagulation quality of warfarin and the role of physician-pharmacist collaborative clinics in the treatment of patients receiving warfarin: a retrospective, observational, single-center study. Front Pharmacol 2021 Jan 14;11:605353 [FREE Full text] [doi: 10.3389/fphar.2020.605353] [Medline: 33519468]
7. Jawitz OK, Wang TY, Lopes RD, Chavez A, Boyer B, Kim H, et al. Rationale and design of PROACT Xa: A randomized, multicenter, open-label, clinical trial to evaluate the efficacy and safety of apixaban versus warfarin in patients with a mechanical On-X Aortic Heart Valve. Am Heart J 2020 Sep;227:91-99 [FREE Full text] [doi: 10.1016/j.ahj.2020.06.014] [Medline: 32693197]
8. Schaefer JK, Li Y, Gu X, Souphis NM, Haymart B, Kline-Rogers E, et al. Association of adding aspirin to warfarin therapy without an apparent indication with bleeding and other adverse events. JAMA Intern Med 2019 Apr 01;179(4):533-541 [FREE Full text] [doi: 10.1001/jamainternmed.2018.7816] [Medline: 30830172]
9. Wallentin L, Yusuf S, Ezekowitz MD, Alings M, Flather M, Franzosi MG, RE-LY investigators. Efficacy and safety of dabigatran compared with warfarin at different levels of international normalised ratio control for stroke prevention in atrial fibrillation: an analysis of the RE-LY trial. Lancet 2010 Sep 18;376(9745):975-983. [doi: 10.1016/S0140-6736(10)61194-4] [Medline: 20801496]
10. Rosendaal FR, Cannegieter SC, van der Meer FJM, Briët E. A method to determine the optimal intensity of oral anticoagulant therapy. Thromb Haemost 1993 Mar 01;69(3):236-239. [Medline: 8470047]
11. Matchar DB, Jacobson A, Dolor R, Edson R, Uyeda L, Phibbs CS, THINRS Executive CommitteeSite Investigators. Effect of home testing of international normalized ratio on clinical events. N Engl J Med 2010 Oct 21;363(17):1608-1620. [doi: 10.1056/NEJMoa1002617] [Medline: 20961244]
12. Menéndez-Jándula B, Souto JC, Oliver A, Montserrat I, Quintana M, Gich I, et al. Comparing self-management of oral anticoagulant therapy with clinic management: a randomized trial. Ann Intern Med 2005 Jan 04;142(1):1-10. [doi: 10.7326/0003-4819-142-1-200501040-00006] [Medline: 15630104]
13. Lo F, Wong EML, Hong FKW. The effects of educational programs on knowledge, international normalized ratio, warfarin adherence, and warfarin-related complications in patients receiving warfarin therapy: an integrative review. J Cardiovasc Nurs 2021 Feb 28:00. [doi: 10.1097/JCN.0000000000000790] [Medline: 33657063]
14. Matchar DB, Samsa GP, Cohen SJ, Oddone EZ, Jurgelski AE. Improving the quality of anticoagulation of patients with atrial fibrillation in managed care organizations: results of the managing anticoagulation services trial. Am J Med 2002 Jul;113(1):42-51. [doi: 10.1016/s0002-9343(02)01131-2] [Medline: 12106622]

15. Yi M, Peng J, Zhang L, Zhang Y. Is the allocation of medical and health resources effective? Characteristic facts from regional heterogeneity in China. Int J Equity Health 2020 Jun 08;19(1):89 [FREE Full text] [doi: 10.1186/s12939-020-01201-8] [Medline: 32513283]
16. Zhu Z, Li Y, Meng X, Han J, Li Y, Liu K, et al. New warfarin anticoagulation management model after heart valve surgery: rationale and design of a prospective, multicentre, randomised trial to compare an internet-based warfarin anticoagulation management model with the traditional warfarin management model. BMJ Open 2019 Dec 05;9(12):e032949 [FREE Full text] [doi: 10.1136/bmjopen-2019-032949] [Medline: 31811010]
17. Grzymala-Lubanski B, Svensson PJ, Renlund H, Jeppsson A, Själander A. Warfarin treatment quality and prognosis in patients with mechanical heart valve prosthesis. Heart 2017 Feb;103(3):198-203. [doi: 10.1136/heartjnl-2016-309585] [Medline: 27590664]
18. Tuckson RV, Edmunds M, Hodgkins ML. Telehealth. N Engl J Med 2017 Oct 19;377(16):1585-1592. [doi: 10.1056/NEJMsr1503323] [Medline: 29045204]
19. Bereznicki LRE, Jackson SL, Peterson GM. Supervised patient self-testing of warfarin therapy using an online system. J Med Internet Res 2013 Jul 12;15(7):e138 [FREE Full text] [doi: 10.2196/jmir.2255] [Medline: 23853350]
20. Prochaska JH, Göbel S, Keller K, Coldewey M, Ullmann A, Lamparter H, et al. Quality of oral anticoagulation with phenprocoumon in regular medical care and its potential for improvement in a telemedicine-based coagulation service--results from the prospective, multi-center, observational cohort study thrombEVAL. BMC Med 2015 Jan 23;13:14 [FREE Full text] [doi: 10.1186/s12916-015-0268-9] [Medline: 25616558]
21. Ryan F, Byrne S, O'Shea S. Randomized controlled trial of supervised patient self-testing of warfarin therapy using an internet-based expert system. J Thromb Haemost 2009 Aug;7(8):1284-1290 [FREE Full text] [doi: 10.1111/j.1538-7836.2009.03497.x] [Medline: 19496921]
22. Zhu Z, Zhang H, Li Y, Meng X, Jiao Y, Qin Y, et al. Establishment of a new platform for the management of patients after cardiac surgery: descriptive study. JMIR Med Inform 2019 Apr 01;7(2):e13123 [FREE Full text] [doi: 10.2196/13123] [Medline: 30932865]
23. Haibo Z, Jinzhong L, Yan L, Xu M. Low-intensity international normalized ratio (INR) oral anticoagulant therapy in Chinese patients with mechanical heart valve prostheses. Cell Biochem Biophys 2012 Jan;62(1):147-151. [doi: 10.1007/s12013-011-9275-4] [Medline: 21892782]
24. Cheung YF, Leung MP. Low dose oral anticoagulation therapy in Chinese children with congenital heart disease. J Paediatr Child Health 1998 Dec;34(6):563-567. [doi: 10.1046/j.1440-1754.1998.00310.x] [Medline: 9928651]
25. Chenhsu R, Chiang S, Chou M, Lin M. Long-term treatment with warfarin in Chinese population. Ann Pharmacother 2000 Dec;34(12):1395-1401. [doi: 10.1345/aph.19289] [Medline: 11144695]
26. You JHS, Chan FWH, Wong RSM, Cheng G. Is INR between 2.0 and 3.0 the optimal level for Chinese patients on warfarin therapy for moderate-intensity anticoagulation? Br J Clin Pharmacol 2005 May;59(5):582-587 [FREE Full text] [doi: 10.1111/j.1365-2125.2005.02361.x] [Medline: 15842557]
27. Chen J, Shao L, Gong L, Luo F, Wang J, Shi Y, et al. A pharmacogenetics-based warfarin maintenance dosing algorithm from Northern Chinese patients. PLoS ONE 2014 Aug 15;9(8):e105250. [doi: 10.1371/journal.pone.0105250]
28. Liu R, Zhang K, Gong Z, Shi X, Zhang Q, Pan X, et al. Association of apolipoprotein E (APOE) polymorphisms with warfarin maintenance dose in a northern Han Chinese population. Lipids Health Dis 2016 Feb 24;15:34 [FREE Full text] [doi: 10.1186/s12944-016-0205-8] [Medline: 26912074]
29. Wang HJ, Ma C, Yang J, Xu Q, Zhang Y, Gao L, et al. CYP4F2 polymorphism as a genetic risk factor for major hemorrhagic complications in Chinese patients on warfarin therapy. J Geriatr Cardiol 2012 Jun;9(2):209-210 [FREE Full text] [doi: 10.3724/SP.J.1263.2012.04162] [Medline: 23097648]
30. Guo Y, Wang H, Tian Y, Wang Y, Lip GY. Time trends of aspirin and warfarin use on stroke and bleeding events in Chinese patients with new-onset atrial fibrillation. Chest 2015 Jul;148(1):62-72 [FREE Full text] [doi: 10.1378/chest.14-2018] [Medline: 25501045]
31. Cao H, Jiang S, Lv M, Wu T, Chen W, Zhang J. Effectiveness of the Alfalfa App in warfarin therapy management for patients undergoing venous thrombosis prevention and treatment: cohort study. JMIR Mhealth Uhealth 2021 Mar 02;9(3):e23332 [FREE Full text] [doi: 10.2196/23332] [Medline: 33650976]
32. Bernstein MR, John L, Sciortino S, Arambages E, Auletta D, Spyropoulos AC. Does telehealth improve anticoagulation management in patient service centers (PSC)? A pilot project. J Thromb Thrombolysis 2020 Feb;49(2):316-320. [doi: 10.1007/s11239-019-02031-4] [Medline: 31898275]
33. Cryder BT, Felczak MA, Darkwa A, Patel H, Janociak JD, Rihani R. Transition of stable patients from traditional anticoagulation clinic services to telephonic management. Int J Clin Pharm 2017 Mar 30;39(3):569-572. [doi: 10.1007/s11096-017-0428-4]
34. Kvedar J, Coye MJ, Everett W. Connected health: a review of technologies and strategies to improve patient care with telemedicine and telehealth. Health Affairs 2014 Feb;33(2):194-199. [doi: 10.1377/hlthaff.2013.0992]
35. Chaet D, Clearfield R, Sabin JE, Skimming K. Ethical practice in telehealth and telemedicine. J GEN INTERN MED 2017 Jun 26;32(10):1136-1140. [doi: 10.1007/s11606-017-4082-2]

36. Stoudenmire LG, DeRemer CE, Elewa H. Telephone versus office-based management of warfarin: impact on international normalized ratios and outcomes. Int J Hematol 2014 Jun 21;100(2):119-124. [doi: 10.1007/s12185-014-1619-6]
37. Xia X, Wu J, Zhang J. The effect of online versus hospital warfarin management on patient outcomes: a systematic review and meta-analysis. Int J Clin Pharm 2018 Dec;40(6):1420-1429. [doi: 10.1007/s11096-018-0731-8] [Medline: 30367374]
38. Boman K, Davidson T, Gustavsson M, Olofsson M, Renström GB, Johansson L. Telemedicine improves the monitoring process in anticoagulant treatment. J Telemed Telecare 2012 Sep;18(6):312-316. [doi: 10.1258/jtt.2012.120319] [Medline: 22892374]

Abbreviations

INR: international normalized ratio
TTR: time in therapeutic range

Edited by R Kukafka; submitted 11.04.21; peer-reviewed by A Bhatt, HH Wang, A Hidki,Dr; comments to author 11.05.21; revised version received 13.05.21; accepted 05.07.21; published 13.08.21.

Please cite as:
Zhu Z, Li C, Shen J, Wu K, Li Y, Liu K, Zhang F, Zhang Z, Li Y, Han J, Qin Y, Yang Y, Fan G, Zhang H, Ding Z, Xu D, Chen Y, Zheng Y, Zheng Z, Meng X, Zhang H
New Internet-Based Warfarin Anticoagulation Management Approach After Mechanical Heart Valve Replacement: Prospective, Multicenter, Randomized Controlled Trial
J Med Internet Res 2021;23(8):e29529
URL: https://www.jmir.org/2021/8/e29529
doi:10.2196/29529
PMID:34397393

©Zhihui Zhu, Chenyu Li, Jinglun Shen, Kaisheng Wu, Yuehuan Li, Kun Liu, Fan Zhang, Zhenhua Zhang, Yan Li, Jie Han, Ying Qin, Yu Yang, Guangpu Fan, Huajun Zhang, Zheng Ding, Dong Xu, Yu Chen, Yingli Zheng, Zhe Zheng, Xu Meng, Haibo Zhang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Internet-Based HIV Self-Testing Among Men Who Have Sex With Men Through Pre-exposure Prophylaxis: 3-Month Prospective Cohort Analysis From China

Jing Zhang[1,2,3,4], PhD; Joseph Tucker[5,6], PhD; Weiming Tang[6,7], PhD; Hongyi Wang[1,2,3,4], PhD; Zhenxing Chu[1,2,3,4], PhD; Qinghai Hu[1,2,3,4], PhD; Xiaojie Huang[8], PhD; Yaokai Chen[9], PhD; Hui Wang[10], PhD; Xiaoqing He[9], MD; Yao Li[9], MD; Lukun Zhang[10], PhD; Zhili Hu[1,2,3,4], MPH; Rantong Bao[1,2,3,4], MPH; Shangcao Li[1,2,3,4], MPH; Hang Li[1,2,3,4], MPH; Haibo Ding[1,2,3,4], PhD; Yongjun Jiang[1,2,3,4], PhD; Wenqing Geng[1,2,3,4], PhD; Junjie Xu[1,2,3,4*], PhD; Hong Shang[1,2,3,4*], PhD

[1]NHC Key Laboratory of AIDS Immunology, National Clinical Research Center for Laboratory Medicine, The First Affiliated Hospital of China Medical University, Shenyang, China

[2]Key Laboratory of AIDS Immunology, Chinese Academy of Medical Sciences, Shenyang, China

[3]Key Laboratory of AIDS Immunology of Liaoning Province, Shenyang, China

[4]Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Zhejiang University, Hangzhou, China

[5]Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom

[6]University of North Carolina at Chapel Hill Project-China, Guangzhou, China

[7]Dermatology Hospital of Southern Medical University, Guangzhou, China

[8]Center for Infectious Diseases, Beijing YouAn Hospital, Capital Medical University, Beijing, China

[9]Division of Infectious Diseases, Chongqing Public Health Medical Center, Chongqing, China

[10]Department of Infectious Diseases, Shenzhen Third People's Hospital, Shenzhen, China

[*]these authors contributed equally

Corresponding Author:
Hong Shang, PhD
NHC Key Laboratory of AIDS Immunology
National Clinical Research Center for Laboratory Medicine
The First Affiliated Hospital of China Medical University
No. 155 Nanjingbei Street
Shenyang, 110001
China
Phone: 86 83282634
Email: hongshang100@hotmail.com

Abstract

Background: Routine HIV testing accompanied with pre-exposure prophylaxis (PrEP) requires innovative support in a real-world setting.

Objective: This study aimed to determine the usage of HIV self-testing (HIVST) kits and their secondary distribution to partners among men who have sex with men (MSM) in China, who use PrEP, in an observational study between 2018 and 2019.

Methods: In 4 major cities in China, we prospectively followed-up MSM from the China Real-world oral PrEP demonstration study, which provides daily or on-demand PrEP for 12 months, to assess the usage and secondary distribution of HIVST on quarterly follow-ups. Half of the PrEP users were randomized to receive 2 HIVSTs per month in addition to quarterly facility-based HIV testing. We evaluated the feasibility of providing HIVST to PrEP users.

Results: We recruited 939 MSM and randomized 471 to receive HIVST, among whom 235 (49.9%) were daily and 236 (50.1%) were on-demand PrEP users. At baseline, the median age was 29 years, 390 (82.0%) men had at least college-level education, and 119 (25.3%) had never undergone facility-based HIV testing before. Three months after PrEP initiation, 341 (74.5%) men had used the HIVST provided to them and found it very easy to use. Among them, 180 of 341 (52.8%) men had distributed the HIVST kits it to other MSM, and 132 (51.6%) among the 256 men who returned HIVST results reported that used it with their sexual partners at the onset of intercourse. Participants on daily PrEP were more likely to use HIVST (adjusted hazard ratio=1.3, 95% CI 1.0-1.6) and distribute HIVST kits (adjusted hazard ratio=1.3, 95% CI 1.1-1.7) than those using on-demand PrEP.

Conclusions: MSM who used PrEP had a high rate of usage and secondary distribution of HIVST kits, especially among those on daily PrEP, which suggested high feasibility and necessity for HIVST after PrEP initiation. Assuming that fourth-generation HIVST kits are available, HIVST may be able to replace facility-based HIV testing to a certain extent.

Trial Registration: Chinese Clinical Trial Registry ChiCTR1800020374; https://www.chictr.org.cn/showprojen.aspx?proj=32481

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2019-036231

(J Med Internet Res 2021;23(8):e23978) doi:10.2196/23978

KEYWORDS

HIV self-testing; men who have sex with men; pre-exposure prophylaxis; secondary distribution; usage

Introduction

Oral pre-exposure prophylaxis (PrEP) can reduce HIV infections among adherent men who have sex with men (MSM) [1-3]. While the adherence to PrEP is only approximately 60% in the real-world setting [4], it is crucial for all PrEP users to undergo quarterly HIV testing to avoid a breakthrough infection and acquiring resistance to antiretroviral therapy [5].

However, the implementation of PrEP has increased the rate of missing facility-based testing. Almost 30% of PrEP users did not show up for their first quarterly clinical visits [6], and approximately 40% of them missed their 6-monthly appointment [7]. Difficulty attending clinical visits and missing laboratory testing has become one of the main reasons for the interruption [8] and discontinuation [9] in PrEP use in clinical settings. Since the outbreak of COVID-19, attendance at facility-based HIV testing became more difficult because of clinic closure, challenges with social distancing, and related difficulties [10]. It is crucial to identify strategies to ensure frequent HIV testing among individuals on PrEP.

In order to improve and expand PrEP uptake, nonclinical PrEP approaches are being piloted, such as central dispensing at pharmacies [11], community venues (eg, automated teller machines in shopping malls) [12], schools, and prisons or jails [13]. These approaches still require periodic HIV testing, often through HIV self-testing (HIVST), and be supplemented with internet-based support, which is a convenient and confidential option for HIV testing, allowing people to take an HIV test and learn the outcome in their own home or at other private locations [14], while mailing and application of the test kit, support consultations during testing, uploading of testing outcomes, and follow-ups can be conducted over the internet [15,16].

HIVST can also promote partner testing through secondary distribution of test kits [17]; that is, an individual who is provided multiple self-test kits can distribute them to sexual partners or to others in their social network [18]. Peers within the key HIV-infected population play an important role in facilitating HIV prevention in that population; for example, by promoting the uptake of HIV testing [19] and linkage to care [20]. Providing multiple HIVST kits to PrEP users and encouraging their secondary distribution would potentially empower MSM using PrEP and the promote HIV prevention within the MSM community.

HIVST technology has advanced in recent years. Third-generation HIVST technology has a testing period of 3 weeks, with high sensitivity and specificity [21]. PrEP recipients can easily administer HIVST at home and know their HIV status at least 4-8 weeks ahead of their next quarterly HIV facility-based test, as recommended by World Health Organization and Centers for Disease Control and Prevention [5]. In such a scenario, HIVST can be used to identify a breakthrough infection earlier and prevent 4-8 weeks of antiretroviral therapy in PrEP that could lead to resistance.

We aimed to assess the usage and secondary distribution of internet-based HIVST kits and their correlates among MSM using PrEP, prospectively. This information can provide evidence regarding the utility of HIVST in the era of PrEP and support a better paired and targeted HIV-testing strategy among PrEP users in the future.

Methods

Study Design, Setting, and Participants

We conducted a randomized control trial of HIV self-testing among MSM who are PrEP users (Chinese Clinical Trial Registry# ChiCTR1800020374). Participants were recruited from among MSM in the ongoing PrEP pragmatic trial (CROPrEP, ChiCTR-IIN-17013762), which provides regimens of daily and on-demand PrEP, in 4 major Chinese cities (Beijing, Shenyang, Chongqing, and Shenzhen) from December 2018 to September 2019. CROPrEP provides all participants standard care of PrEP, including quarterly facility-based HIV testing. Given the well-documented protective effect of PrEP, the sample size of the CROPrEP study at each site was not determined through power calculations. The sample size at each site was increased maximally by fully considering the human resources and capability of each study site [22].

We randomized half of the enrolled PrEP recipients to provide them 2 HIV self-tests per month, in addition to the standard care. The study team has generated a service account on one of the most popular social media in China (WeChat) to provide web-based services on the application of extra testing kits, instructions on self-testing, real-time consultation with the staff, uploading of test outcomes, and follow-up questionnaires. All HIVST kits distributed or secondarily distributed were marked with a serial number. Participants who received HIVST were also encouraged to share them with their male sexual partners.

All participants were prospectively followed up quarterly. This study adhered to the STROBE (Strengthening the Reporting of Observational studies in Epidemiology) checklist (Multimedia Appendix 1).

The study sites are general hospitals equipped with HIV voluntary counseling and testing clinics and HIV treatment clinics with physicians specialized in infectious diseases in 4 major cities in China: The Youan Hospital of Capital Medical University, Beijing; The First Affiliated Hospital of China Medical University, Shenyang; The Chongqing Public Health Medical Center of Southwest University, Chongqing; and The Third People's Hospital, Shenzhen.

Participants were eligible to participate in this study if they (1) were designated male at birth; (2) have sex with male partners; (3) are aged 18-65 years; (4) report having ≥1 of the following risk factors in the last 6 months: (a) had condomless receptive anal sex with a male partner, (b) had more than 2 male sexual partners, (c) had self-reported sexually transmitted infections (STIs) such as syphilis, gonorrhea, chlamydia, chancroid, or lymphogranuloma venereum, or (d) have ever used postexposure prophylaxis medication, but have not received postexposure prophylaxis medication in the previous month; (5) have a nonreactive outcome on a fourth-generation HIV enzyme-linked immunosorbent assay test at baseline screening and undetectable HIV-1 RNA; (6) have no evidence of severe liver or kidney dysfunction on a comprehensive evaluation (including physical examination, urine test, and blood biochemical examination); and (7) indicate willingness to participate and sign an informed consent form. Individuals were excluded if they (1) are deemed ineligible on eligibility evaluation for the CROPrEP trial, (2) refuse to accept or use the HIVST kits (with reasons recorded), or (3) refuse to sign the informed consent form.

To prevent loss to follow-up bias, the clinicians and study staff provided one-on-one personalized compliance support, counseling, and cohort maintenance during the follow-up period. Furthermore, community-based organizations and the weekly internet-based retention strategies were used to strengthen group- and individual-level supervision of retention and cohort management. This internet-based strategy included interactive peer counseling focusing on study retention; a short message providing routine follow-up visit reminders, along with a live chat.

Measures

Information on demographic characteristics (including age, education level, marital status, and monthly income), HIV-related risk behavior in the past 3 months (frequency of anal intercourse, and instances of condomless anal intercourse, substance abuse, etc), and testing history for HIV and STIs in the past were collected at baseline through self-administrated questionnaires distributed on each participants' smartphone. The aforementioned information was obtained in accordance with a systematic review that analyzed studies on HIV testing behaviors among MSM in China [23]. Participants were grouped into <25-year-old and ≥25-year-old age groups in accordance with the definition of youth of the United Nations [24]. Participants were also grouped by their median monthly income. Measures of monthly income were separated by the median income. Usage of HIVST and secondary distribution of HIVST to male sexual partners and other MSM were recorded at quarterly follow-up among all participants through self-administrated questionnaires on each participants' smartphone. The returned results of photographs of used HIVST were uploaded by participants after they used HIVST kits, and information regarding the place, occasion, and ease of using HIVST was collected. The proportion of HIVST usage was calculated among participants retained at the third month follow-up visit. Furthermore, the proportion of HIVST secondary distribution was calculated among participants who had self-reported usage of HIVST at the third month follow-up visit.

Statistical Analyses

Statistical analysis was conducted using SPSS (version 20.0, IBM Corp). Demographics, the PrEP regimen, and HIV-related risk behavior in the past 3 months among all participants were expressed as numbers and percentages. Follow-up time was determined as the number of days from the date of enrollment to the date of quarterly follow-up. Univariable and multivariable Cox regression analyses were performed to assess the predictors of self-reported usage of HIVST and secondary distribution of HIVST at the first quarterly follow-up. Variables with P<.20 on univariable analysis were included in the multivariable model to avoid the omission of clinically relevant variables, which had an underestimated effect in univariable analysis [25,26]. Variables in the final model were selected with a forward stepwise procedure. Hazard ratios (HRs) and adjusted HRs (aHRs) were calculated. On multivariate analysis, P<.05 was considered the cut-off for a significant difference. We plotted Kaplan–Meier survival curves for predictors for HIVST usage and their secondary distribution during multivariable Cox regression analysis with P<.05 indicating significance.

Ethics Approval

The study protocol was approved by the ethics review board of the First Affiliated Hospital of China Medical University (IRB-2018-273), Shenyang. Written informed consent was obtained from each participant before collecting study information or blood samples. Participants voluntarily participated in the study and had the right to refuse to answer any question. Participants had the right to withdraw from the study without penalty. The protocols for the CROPrEP trial [25] and this study [27] have been published. This HIVST study among PrEP recipients was registered on the Chinese Clinical Trial Registry (trial ID ChiCTR1800020374).

Results

Baseline Demographics and Behavioral Characteristics

A total of 939 MSM were recruited from the CROPrEP trial, with 470 daily PrEP users and 469 on-demand PrEP users. In total, 471 men were randomized to receive 2 HIV self-tests per month, in addition to the standard care, which included 235 (49.0%) daily PrEP users and 236 (50.1%) on-demand PrEP users.

The median age of these 471 participants at baseline was 29 years (quantile 25-35 years); among them, 390 (82.0%) had an education level of undergraduate and above, 241 (51.2%) had a monthly income of less than US $857, and 266 (56.5%) were single. Regarding HIV-related risk behavior in the past 3 months, 165 (35.0%) individuals had had anal intercourse every week and 234 (49.7%) had had anal intercourse every month, 312 (66.2%) of them had condomless anal intercourse with male sexual partners, 226 (48.0%) used Poppers in the previous 3 months, and 103 (22.5%) had participated in group sex. In total, 39 (8.3%) of them self-reported having STI-related symptoms in the previous 12 months. Further, 375 (79.6%) participants self-reported having used HIVST in the past, and 352 (74.7%) reported having undergone facility-based HIV testing in the past (Table 1).

Table 1. Baseline characteristics of pre-exposure prophylaxis recipients carrying out HIV self-testing in 4 major cities of China (N=471).

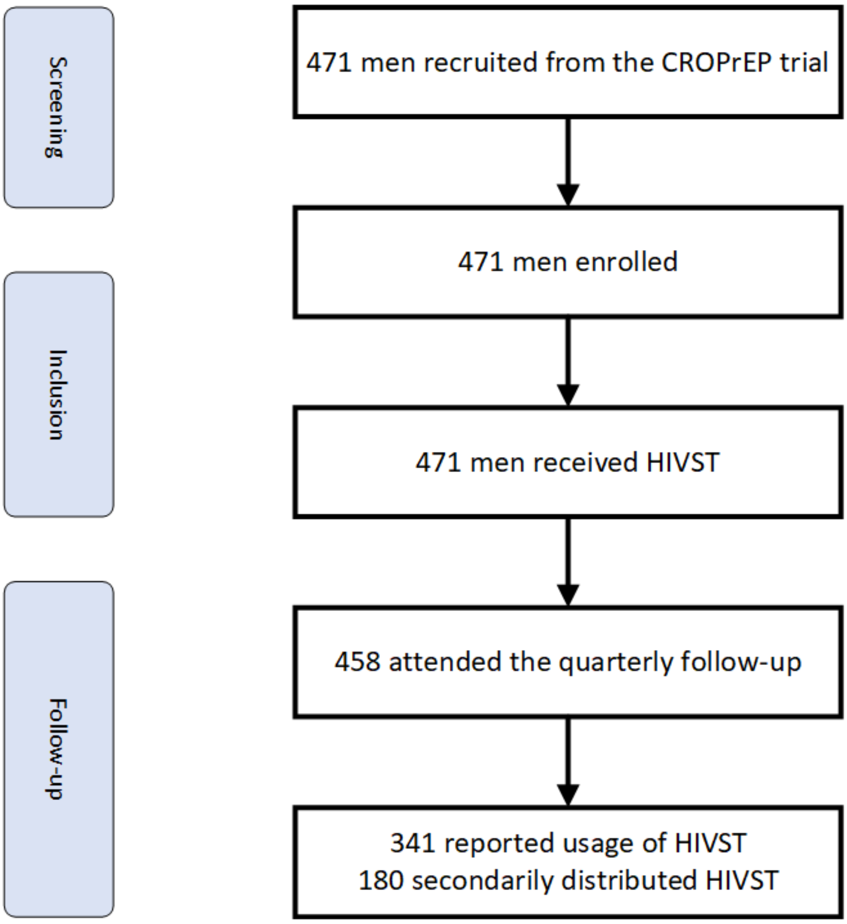
Variable	Participants, n (%)
Pre-exposure prophylaxis regimen	
Daily	235 (49.9)
On-demand	236 (50.1)
Age (median 29.0 years, quantiles 25.0-35.0 years)	
<25	84 (17.8)
≥25	387 (82.2)
Education level	
Senior high and below	81 (17.2)
Undergraduate and above	390 (82.8)
Monthly income (US $)	
≤857	241 (51.2)
858	230 (48.8)
Marital status	
Single	266 (56.5)
Married or cohabiting with a female	30 (6.4)
In a relationship or cohabitating with a male	175 (37.2)
Frequency of anal intercourse in the past 3 months	
Every day	8 (1.7)
Every week	165 (35.0)
Every month	234 (49.7)
Less than once a month	64 (13.6)
Had condomless anal intercourse in the previous 3 months	
Yes	312 (66.2)
No	159 (33.8)
Had used Poppers in the previous 3 months	
Yes	226 (48.0)
No	245 (52.0)
Had group sex in the previous 3 months[a]	
Yes	103 (22.5)
No	354 (77.5)
Had sexually transmitted infection–related symptoms in the previous 12 months	
Yes	39 (8.3)
No	432 (91.7)
Have ever used HIV self-testing in the past	
Yes	375 (79.6)
No	96 (20.4)
Have ever undergone facility-based HIV testing in the past	
Yes	352 (74.7)
No	119 (25.3)

[a] There were 14 missing data.

Usage, Returning Results, and Sharing of HIVST at the First Quarterly Follow-ups

Among these 471 men, 458 men were retained at the first quarterly follow-up, resulting in a retention rate of 97.2%. At the first quarterly of PrEP initiation, a total of 341 (74.5%) participants self-reported that they had used the HIVST in the past 3 months, among whom 256 (75.1%) returned HIVST testing outcomes by uploading photographs of the used HIVST kits marked with a serial number. In total, 260 (76.2%) men self-reported that they had recommended HIVST to their male sexual partners or gay friends, and 180 (52.8%) participants self-reported that they had shared their HIVST kits with the male sexual partners or gay friends. Among the 256 participants who had returned HIVST outcomes, the most frequent place of using HIVST was at home (n=218, 85.2%). In total, 132 (51.6%) men reported that HIVST was used immediately before or after sexual intercourse with their male sexual partners (Figure 1).

Figure 1. Study flow diagram. HIVST: HIV self-testing.


Screening
Inclusion
Follow-up
471 men recruited from the CROPrEP trial
471 men enrolled
471 men received HIVST
458 attended the quarterly follow-up
341 reported usage of HIVST
180 secondarily distributed HIVST

Correlation of Self-Reported Usage of HIVST on Quarterly Follow-up

On univariable Cox regression analysis, men with the following variables were more likely to use HIVST during follow-up: being on daily PrEP (vs on-demand PrEP, HR=1.273, 95% CI 1.028-1.576; *P*=.03), having an education level of senior high or below (vs college and above, HR=1.393, 95% CI 1.054-1.841; *P*=.02), having STI-related symptoms in the previous 12 months (vs not having symptoms, HR=1.384, 95% CI 0.956-2.004; *P*=.09), having ever used HIVST in the past (vs having used HIVST in the past, HR=1.492, 95% CI 1.102-2.020; *P*=.01). Participants with the following variables were less likely to use HIVST: having had anal intercourse every day in the past 3 months (vs less than once a month, HR=0.396, 95% CI 0.143-1.101; *P*=.08), having had condomless anal intercourse in the past 3 months (vs not having had condomless anal intercourse, HR=0.825, 95% CI 0.687-0.990; *P*=.04) (Table 2).

After the forward stepwise procedure, all variables with *P*<.20 on univariable regression analysis were included in the multivariable regression model, and participants with the following variables were more likely to use HIVST during follow-ups: being on daily PrEP (vs on-demand PrEP, aHR=1.298, 95% CI 1.047-1.608; *P*=.02), having an education level of senior high or below (vs undergraduate and above, aHR=1.482, 95% CI 1.119-1.964; *P*=.006), and having used HIVST in the past (vs not having used HIVST in the past, aHR=1.642, 95% CI 1.205-2.242; *P*=.002). Furthermore, participants who self-reported having used Poppers in the past 3 months (vs those not having used Poppers in the past 3 months, aHR=0.788, 95% CI 0.633-9.981; *P*=.03) were less likely to use HIVST (Table 2 and Figure 2).

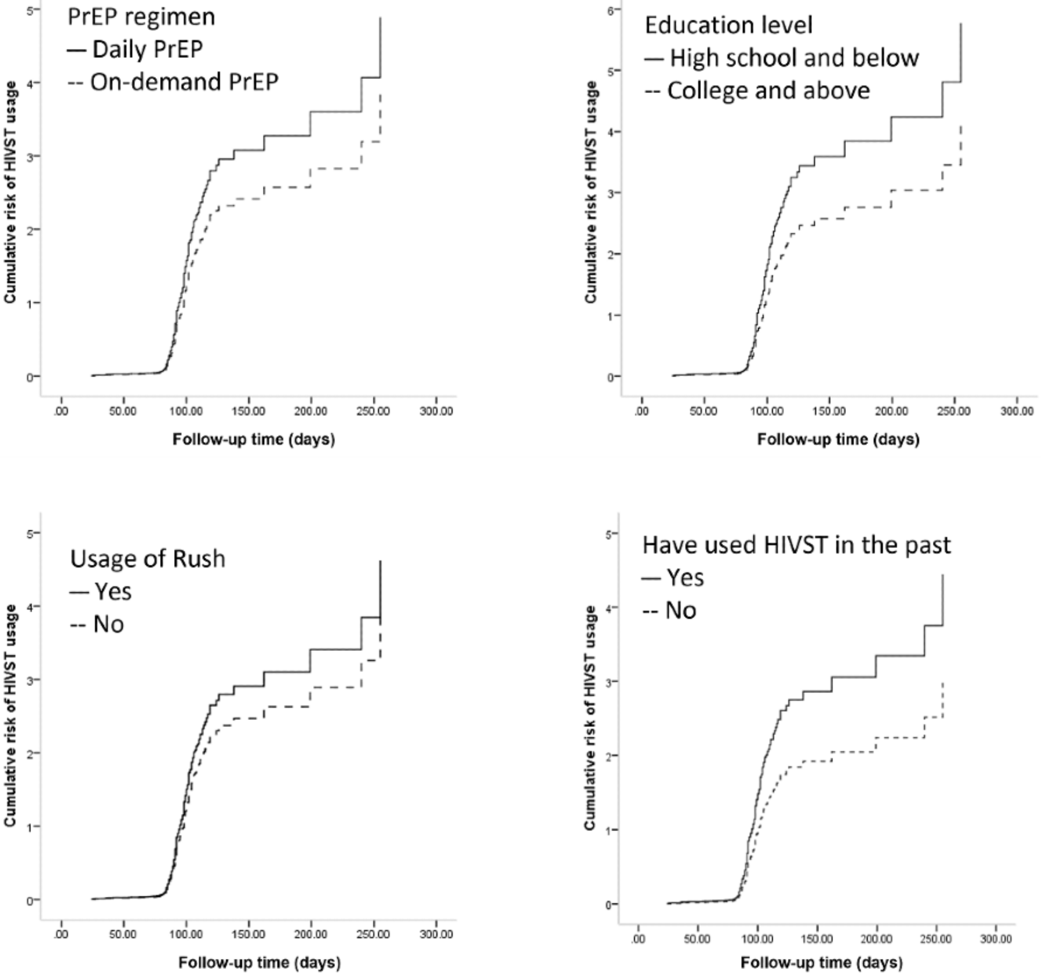
Table 2. Cox regression model of correlates of usage of HIV self-testing among pre-exposure prophylaxis recipients within 3 months of treatment initiation in 4 major cities in China (N=471).

Variable	Usage of HIV self-testing				Secondary distribution of HIV self-test kits			
	Hazard ratio (95% CI)	*P* value	Adjusted hazard ratio (95% CI)[a]	*P* value	Hazard ratio (95% CI)	*P* value	Adjusted hazard ratio (95% CI)	*P* value
Pre-exposure prophylaxis regimen of daily (vs on-demand) users	1.273 (1.028-1.576)	.03	1.298 (1.047-1.608)	.02	1.343 (1.072-1.682)	.01	1.352 (1.079-1.694)	.009
Age<25 years	0.990 (0.751-1.307)	.95	—[b]	—	1.066 (0.739-1.539)	.73	—	—
Education level of high school and below	1.393 (1.054-1.841)	.02	1.482 (1.119-1.964)	.006	1.151 (0.840-1.578)	.38	—	—
Monthly income <US $857	1.149 (0.928-1.423)	.20	—	—	1.074 (0.858-1.346)	.53	—	—
Currently married (vs single)	1.436 (0.811-2.263)	.12	—	—	1.523 (0.963-2.408)	.07	1.706 (1.074-2.709)	.02
Frequency of anal intercourse in the previous 3 months								
Every day	1.0	Reference	—	—	1.0	Reference	—	—
Every week	1.969 (0.726-5.342)	.18	—	—	1.480 (0.603-3.634)	.39	—	—
Every month	2.102 (0.779-5.673)	.14	—	—	1.511 (0.619-3.688)	.36	—	—
Less than once a month	2.523 (0.909-7.004)	.08	—	—	1.481 (0.580-3.783)	.41	—	—
Had condomless anal intercourse in the past 3 months	0.825 (0.687-0.990)	.04	—	—	—	.33	—	—
Used Poppers in the past 3 months	0.848 (0.684-1.049)	.13	0.788 (0.633-0.981)	.03	0.953 (0.761-1.193)	.68	—	—
Had group sex in the past 3 months	1.011 (0.781-1.310)	.93	—	—	0.948 (0.724-1.241)	.70	—	—
Had sexually transmitted infection–related symptoms in the past 12 months	1.384 (0.956-2.004)	.09	—	—	1.154 (0.760-1.753)	.50	—	—
Have performed HIV self-testing in the past	1.492 (1.102-2.020)	.01	1.643 (1.205-2.242)	.002	1.415 (1.034-1.937)	.03	1.464 (1.066-2.009)	.02
Have undergone facility-based HIV testing in the past	1.166 (0.904-1.503)	.24	—	—	1.131 (0.867-1.476)	.36	—	—

[a] Variables with *P*<.20 on univariable analysis were included in the multivariable Cox regression model.

[b] —: not determined.

Figure 2. Kaplan–Meier survival curves for predictors of HIVST usage. HIVST: HIV self-testing, PrEP: pre-exposure prophylaxis.

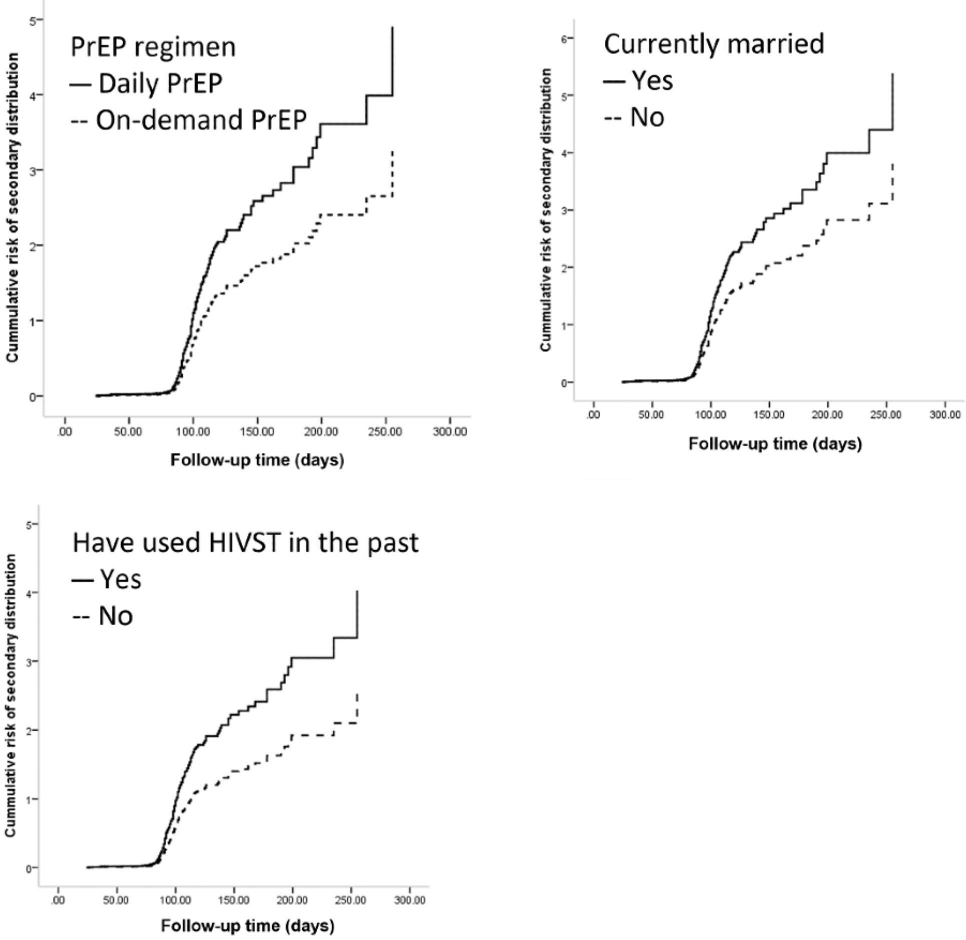


Correlation of Secondary Distribution of HIVST Kits at Quarterly Follow-up

On univariable Cox regression analysis, participants with the following variables were more likely to secondarily distribute HIVST kits during follow-up: being on daily PrEP (vs on-demand PrEP, HR=1.343, 95% CI 1.072-1.682; *P*=.01), being currently married (vs being single, HR=1.523, 95% CI 0.963-2.408; *P*=.07), having ever used HIVST in the past (vs having never used HIVST in the past, HR=1.415, 95% CI 1.034-1.937; *P*=.03) (Table 2).

After the stepwise forward procedure, all variables with *P*<.20 on univariable regression were included in the multivariable regression model, and participants with following variables were more likely to secondarily distribute HIVST during follow-ups: being on daily PrEP (vs on-demand PrEP, aHR=1.352, 95% CI 1.079-1.694; *P*=.009), being currently married (vs single, aHR=1.706, 95% CI 1.074-2.709; *P*=.02), and having used HIVST in the past (vs not having used HIVST in the past, aHR=1.464, 95% CI 1.066-2.009 *P*=.02) (Figure 3).

Figure 3. Kaplan–Meier survival curves for predictors of HIVST secondary distribution. HIVST: HIV self-testing; PrEP: pre-exposure prophylaxis.



Discussion

Principal Findings

We prospectively identified a high rate of usage and secondary distribution of the internet-based HIVST among MSM who are using PrEP in a multicenter pragmatic trial in China, especially among the daily PrEP users. This result warrants the internet-based HIVST as a crucial part of the HIV testing strategy paired with PrEP among the MSM population. Furthermore, the high rate of secondary distribution of HIVST kits and self-testing among sexual partners among PrEP users assures a promising future for peer-initiated expanded HIV testing and maximization of the HIV prevention among the MSM community.

Approximately 80% of PrEP users have used HIVST within 3 months after PrEP initiation and found it easy to use, reflecting HIV testing behavior among PrEP recipients. Every 4 of 5 of these PrEP users had used HIVST before the initiation of PrEP, which is much higher than the level among the general MSM population in China (20%-40%) [28,29]. Since difficulty attending facility-based testing was one of the main reasons for discontinuing PrEP during implementation [8,9], this high usage of HIVST among PrEP users warrants future consideration of substituting some facility-based HIV testing during PrEP and provides evidence in support of HIVST to support PrEP users during the COVID-19 pandemic in the future, since facility-based services and in-person patient-clinician contact has been limited because of the pandemic [30].

Notably, there was a high rate of secondary distribution of HIVST and testing of sexual partners among those who had used HIVST after PrEP initiation. Half of the HIVST was used to test together with their sexual partners immediately before or after intercourse. This is significant because of the limited success and high cost of accessing MSM at clinics or provider-initiated expanded HIV testing strategies [31,32]. According to previous studies, MSM who accept or uptake PrEP were more concerned about their partners' HIV status [33] and had more sexual partners [34], which makes them excellent facilitators for expanded HIV testing among the MSM population. Empowering them with HIVST as a tool to undertake secondary distribution and partner testing can maximize the impact of PrEP on the MSM community via peer-initiated HIV prevention campaigns. Future studies should vigorously pursue and explore the critical role of PrEP users in HIV prevention within the key populations.

Factors promoting the usage and secondary distribution of HIVST among PrEP users were also identified. Men on daily PrEP were more likely to use and secondarily distribute HIVST after PrEP initiation than those on the on-demand regimen. Since most of the HIVST was used immediately before or after sexual intercourse, this correlation can be explained by the fact that MSM on daily PrEP displayed more frequent HIV-related risk behavior and a higher number of sexual partners [35,36].

According to HIV transmission network studies, MSM with a higher number of sexual partners are associated more firmly with other MSM and had a crucial impact on the network of HIV transmission within the population. This correlation illustrates that MSM on daily PrEP play a more critical role in expanded HIV testing and HIV prevention campaigns as peers among the MSM community. Studies on the social network of PrEP recipients built through HIVST distribution should be actively pursued in the future.

An inhibitor of the use of HIVST among PrEP users is the use of Poppers, which is one of the most popular recreational drugs among MSM worldwide [37]. According to previous studies, Poppers users are more likely to consider themselves PrEP candidates and more likely to be a current PrEP user [38]. However, Poppers usage among MSM is associated with an increased risk of HIV infection [39,40] and a higher probability of participating in group sex [40], which indicates the necessity of timely HIV testing. This inhibitor should draw attention from PrEP providers to address timely HIV testing for PrEP users.

Limitations

This study has several limitations. First, there is no fourth-generation HIVST kit available currently in China. These are available in the United Kingdom and in other countries and would potentially allow this model to replace some of the facility-based testing. Second, this study was conducted before the COVID-19 pandemic, which might underestimate the usage of HIVST among PrEP users. Third, this study only assessed the usage and secondary distribution within the first 3 months after PrEP initiation, which requires a longer follow-up period to illustrate its long-term usage. Finally, this study only included participants from 4 major cities in China, which did not represent the general MSM population in China.

Conclusions

Within a pragmatic setting, MSM who are using PrEP had a high rate of usage and secondary distribution of HIVST kits, especially among those on daily PrEP. Offering multiple HIVST to PrEP users to replace some of the facility-based HIV testing and to facilitate expanded HIV testing is feasible and necessary, especially during the COVID-19 pandemic.

Acknowledgments

The authors would like to thank the community-based organizations of *Sunshine*, Shenyang; *YouAn-Tian Yuan Studio*, Beijing; *Chongqing Red Ribbon AIDS Care Center*, Chongqing; and *258 Rainbow Workshop*; Shenzhen, for their help with recruiting and maintaining the participants. We would also like to thank all of the participants for being involved in our study. This work was supported by the Mega-Projects of national science research (13rd Five-Year Plan [2017ZX10201101-002-007]), the Mega-Projects of national science research (12th Five-Year Plan [2012ZX10001006-001-010]), National Natural Science Foundation of China (81872674), and Mega-Projects of national science research (2018ZX10101001-001-003).

Authors' Contributions

JZ, JJX, and HS conceived and designed the study. JZ, HYW, ZXC, QHH, XJH, YKC, HW, XQH, YL, LKZ, ZLH, RTB, SCL, HL, HBD, YJJ, and WQG performed the study and experiments. JZ, JJX, JDT, and WMT drafted the manuscript. All authors reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
STROBE checklist.
[PDF File (Adobe PDF File), 102 KB - jmir_v23i8e23978_app1.pdf]

References

1. Grant RM, Lama JR, Anderson PL, McMahan V, Liu AY, Vargas L, iPrEx Study Team. Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. N Engl J Med 2010 Dec 30;363(27):2587-2599 [FREE Full text] [doi: 10.1056/NEJMoa1011205] [Medline: 21091279]
2. Anderson PL, Glidden DV, Liu A, Buchbinder S, Lama JR, Guanira JV, iPrEx Study Team. Emtricitabine-tenofovir concentrations and pre-exposure prophylaxis efficacy in men who have sex with men. Sci Transl Med 2012 Sep 12;4(151):151ra125 [FREE Full text] [doi: 10.1126/scitranslmed.3004006] [Medline: 22972843]
3. McCormack S, Dunn DT, Desai M, Dolling DI, Gafos M, Gilson R, et al. Pre-exposure prophylaxis to prevent the acquisition of HIV-1 infection (PROUD): effectiveness results from the pilot phase of a pragmatic open-label randomised trial. Lancet 2016 Jan 02;387(10013):53-60 [FREE Full text] [doi: 10.1016/S0140-6736(15)00056-2] [Medline: 26364263]
4. Yun K, Xu J, Zhang J, Li J, Hu Q, Chu Z, et al. Female and younger subjects have lower adherence in PrEP trials: a meta-analysis with implications for the uptake of PrEP service to prevent HIV. Sex Transm Infect 2018 May;94(3):163-168. [doi: 10.1136/sextrans-2017-053217] [Medline: 28756409]

5. What's New in the Preexposure Prophylaxis for the Prevention of HIV Infection in the United States - 2017 Update – A Clinical Practice Guideline? Centers for Disease Control and Prevention. URL: https://www.cdc.gov/hiv/pdf/risk/prep/cdc-hiv-prep-guidelines-2017.pdf [accessed 2021-07-23]
6. Parsons JT, Rendina HJ, Lassiter JM, Whitfield THF, Starks TJ, Grov C. Uptake of HIV Pre-Exposure Prophylaxis (PrEP) in a National Cohort of Gay and Bisexual Men in the United States. J Acquir Immune Defic Syndr 2017 Mar 01;74(3):285-292 [FREE Full text] [doi: 10.1097/QAI.0000000000001251] [Medline: 28187084]
7. Hevey MA, Walsh JL, Petroll AE. PrEP Continuation, HIV and STI Testing Rates, and Delivery of Preventive Care in a Clinic-Based Cohort. AIDS Educ Prev 2018 Oct;30(5):393-405 [FREE Full text] [doi: 10.1521/aeap.2018.30.5.393] [Medline: 30332309]
8. Scott HM, Spinelli M, Vittinghoff E, Morehead-Gee A, Hirozawa A, James C, et al. Racial/ethnic and HIV risk category disparities in preexposure prophylaxis discontinuation among patients in publicly funded primary care clinics. AIDS 2019 Nov 15;33(14):2189-2195 [FREE Full text] [doi: 10.1097/QAD.0000000000002347] [Medline: 31436610]
9. Rolle C, Onwubiko U, Jo J, Sheth AN, Kelley CF, Holland DP. PrEP Implementation and Persistence in a County Health Department Setting in Atlanta, GA. AIDS Behav 2019 Oct;23(Suppl 3):296-303 [FREE Full text] [doi: 10.1007/s10461-019-02654-x] [Medline: 31468296]
10. Jiang H, Zhou Y, Tang W. Maintaining HIV care during the COVID-19 pandemic. Lancet HIV 2020 May;7(5):e308-e309 [FREE Full text] [doi: 10.1016/S2352-3018(20)30105-3] [Medline: 32272084]
11. Tung EL, Thomas A, Eichner A, Shalit P. Implementation of a community pharmacy-based pre-exposure prophylaxis service: a novel model for pre-exposure prophylaxis care. Sex Health 2018 Nov;15(6):556-561. [doi: 10.1071/SH18084] [Medline: 30401342]
12. Stankevitz K, Schwartz K, Hoke T, Li Y, Lanham M, Mahaka I, et al. Reaching at-risk women for PrEP delivery: What can we learn from clinical trials in sub-Saharan Africa? PLoS One 2019;14(6):e0218556 [FREE Full text] [doi: 10.1371/journal.pone.0218556] [Medline: 31226165]
13. Venter WDF. Pre-exposure Prophylaxis: The Delivery Challenge. Front Public Health 2018;6:188 [FREE Full text] [doi: 10.3389/fpubh.2018.00188] [Medline: 30035106]
14. Self-Testing. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/hiv/testing/self-testing.html [accessed 2020-04-23]
15. Ngure K, Heffron R, Mugo N, Thomson KA, Irungu E, Njuguna N, et al. Feasibility and acceptability of HIV self-testing among pre-exposure prophylaxis users in Kenya. J Int AIDS Soc 2017 Feb 10;20(1):21234 [FREE Full text] [doi: 10.7448/IAS.20.1.21234] [Medline: 28362073]
16. Ortblad KF, Chanda MM, Musoke DK, Ngabirano T, Mwale M, Nakitende A, et al. Acceptability of HIV self-testing to support pre-exposure prophylaxis among female sex workers in Uganda and Zambia: results from two randomized controlled trials. BMC Infect Dis 2018 Oct 04;18(1):503 [FREE Full text] [doi: 10.1186/s12879-018-3415-z] [Medline: 30286737]
17. Masters SH, Agot K, Obonyo B, Napierala Mavedzenge S, Maman S, Thirumurthy H. Promoting Partner Testing and Couples Testing through Secondary Distribution of HIV Self-Tests: A Randomized Clinical Trial. PLoS Med 2016 Nov;13(11):e1002166 [FREE Full text] [doi: 10.1371/journal.pmed.1002166] [Medline: 27824882]
18. Thirumurthy H, Masters SH, Mavedzenge SN, Maman S, Omanga E, Agot K. Promoting male partner HIV testing and safer sexual decision making through secondary distribution of self-tests by HIV-negative female sex workers and women receiving antenatal and post-partum care in Kenya: a cohort study. Lancet HIV 2016 Jun;3(6):e266-e274 [FREE Full text] [doi: 10.1016/S2352-3018(16)00041-2] [Medline: 27240789]
19. Choko AT, Nanfuka M, Birungi J, Taasi G, Kisembo P, Helleringer S. A pilot trial of the peer-based distribution of HIV self-test kits among fishermen in Bulisa, Uganda. PLoS One 2018;13(11):e0208191 [FREE Full text] [doi: 10.1371/journal.pone.0208191] [Medline: 30496260]
20. Chang LW, Nakigozi G, Billioux VG, Gray RH, Serwadda D, Quinn TC, et al. Effectiveness of peer support on care engagement and preventive care intervention utilization among pre-antiretroviral therapy, HIV-infected adults in Rakai, Uganda: a randomized trial. AIDS Behav 2015 Oct;19(10):1742-1751 [FREE Full text] [doi: 10.1007/s10461-015-1159-y] [Medline: 26271815]
21. Figueroa C, Johnson C, Ford N, Sands A, Dalal S, Meurant R, et al. Reliability of HIV rapid diagnostic tests for self-testing compared with testing by health-care workers: a systematic review and meta-analysis. Lancet HIV 2018 Jun;5(6):e277-e290 [FREE Full text] [doi: 10.1016/S2352-3018(18)30044-4] [Medline: 29703707]
22. Wang H, Zhang Y, Mei Z, Jia Y, Leuba SI, Zhang J, et al. Protocol for a multicenter, real-world study of HIV pre-exposure prophylaxis among men who have sex with men in China (CROPrEP). BMC Infect Dis 2019 Aug 15;19(1):721 [FREE Full text] [doi: 10.1186/s12879-019-4355-y] [Medline: 31416439]
23. Zou H, Hu N, Xin Q, Beck J. HIV testing among men who have sex with men in China: a systematic review and meta-analysis. AIDS Behav 2012 Oct;16(7):1717-1728. [doi: 10.1007/s10461-012-0225-y] [Medline: 22677975]
24. Youth. United Nations. URL: https://www.un.org/en/sections/issues-depth/youth-0/index.html [accessed 2020-12-31]
25. Kang S, Cho Y, Park G, Ahn J, Han S, Lee J, et al. Predictors for functionally significant in-stent restenosis: an integrated analysis using coronary angiography, IVUS, and myocardial perfusion imaging. JACC Cardiovasc Imaging 2013 Nov;6(11):1183-1190 [FREE Full text] [doi: 10.1016/j.jcmg.2013.09.006] [Medline: 24229771]

26. Zhang J, Xu J, Leuba SI, Chu Z, Hu Q, Mao X, et al. Increasing Condomless Anal Intercourse and HIV Prevalence Among Men Who Have Sex with Men Who Have Never Been Tested for HIV Before: A Serial Cross-Sectional Study in Shenyang, China from 2012 to 2016. AIDS Behav 2020 Oct;24(10):2918-2926. [doi: 10.1007/s10461-020-02842-0] [Medline: 32297067]
27. Zhang J, Huang X, Chen Y, Wang H, Zhang Y, Wang H, CROPrEP Group. HIV and other STIs self-testing to reduce risk compensation among men who have sex with men who use oral pre-exposure prophylaxis in China: protocol for a randomised waitlist-controlled trial. BMJ Open 2020 Jul 19;10(7):e036231 [FREE Full text] [doi: 10.1136/bmjopen-2019-036231] [Medline: 32690516]
28. Han L, Bien CH, Wei C, Muessig KE, Yang M, Liu F, et al. HIV self-testing among online MSM in China: implications for expanding HIV testing among key populations. J Acquir Immune Defic Syndr 2014 Oct 01;67(2):216-221 [FREE Full text] [doi: 10.1097/QAI.0000000000000278] [Medline: 24991972]
29. Ren X, Wu Z, Mi G, McGoogan JM, Rou K, Zhao Y, et al. HIV care-seeking behaviour after HIV self-testing among men who have sex with men in Beijing, China: a cross-sectional study. Infect Dis Poverty 2017 Jun 28;6(1):112 [FREE Full text] [doi: 10.1186/s40249-017-0326-y] [Medline: 28655340]
30. Self-Testing. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/hiv/testing/self-testing.html [accessed 2020-06-01]
31. Kennedy CE, Fonner VA, Sweat MD, Okero FA, Baggaley R, O'Reilly KR. Provider-initiated HIV testing and counseling in low- and middle-income countries: a systematic review. AIDS Behav 2013 Jun;17(5):1571-1590 [FREE Full text] [doi: 10.1007/s10461-012-0241-y] [Medline: 22752501]
32. Leon N, Mathews C, Lewin S, Osler M, Boulle A, Lombard C. A comparison of linkage to HIV care after provider-initiated HIV testing and counselling (PITC) versus voluntary HIV counselling and testing (VCT) for patients with sexually transmitted infections in Cape Town, South Africa. BMC Health Serv Res 2014 Aug 18;14:350 [FREE Full text] [doi: 10.1186/1472-6963-14-350] [Medline: 25134822]
33. Kwan TH, Lee SS. Bridging Awareness and Acceptance of Pre-Exposure Prophylaxis Among Men Who Have Sex With Men and the Need for Targeting Chemsex and HIV Testing: Cross-Sectional Survey. JMIR Public Health Surveill 2019 Jul 03;5(3):e13083 [FREE Full text] [doi: 10.2196/13083] [Medline: 31271148]
34. Brooks RA, Landrian A, Lazalde G, Galvan FH, Liu H, Chen Y. Predictors of Awareness, Accessibility and Acceptability of Pre-exposure Prophylaxis (PrEP) Among English- and Spanish-Speaking Latino Men Who have Sex with Men in Los Angeles, California. J Immigr Minor Health 2020 Aug;22(4):708-716. [doi: 10.1007/s10903-019-00955-w] [Medline: 31823164]
35. Molina JM, Ghosn J, Béniguel L, Rojas-Castro D, Algarte-Genin M, Pialoux G, Prevenir ANRS study group. Incidence of HIV-infection in the ANRS Prevenir study in Paris region with daily or on-demand PrEP with TDF/FTC. 2018 Jul 26 Presented at: 22nd International AIDS Conference; July 23-27, 2018; Amsterdam URL: http://programme.aids2018.org/Abstract/Abstract/13278, [doi: 10.1002/jia2.25148]
36. Zimmermann HM, Eekman SW, Achterbergh RC, Schim van der Loeff MF, Prins M, de Vries HJC, Amsterdam PrEP Project team in the HIV Transmission Elimination AMsterdam Consortium (H-TEAM). Motives for choosing, switching and stopping daily or event-driven pre-exposure prophylaxis - a qualitative analysis. J Int AIDS Soc 2019 Oct;22(10):e25389 [FREE Full text] [doi: 10.1002/jia2.25389] [Medline: 31612621]
37. Bae K, Kwon NJ, Han E. A review on the abuse of three NPS (synthetic cannabinoids, kratom, poppers) among youths in Asia. Forensic Sci Int 2018 Nov;292:45-49. [doi: 10.1016/j.forsciint.2018.09.008] [Medline: 30273913]
38. Hambrick HR, Park SH, Schneider JA, Mayer KH, Carrico AW, Sherman SE, et al. Poppers and PrEP: Use of Pre-exposure Prophylaxis Among Men Who Have Sex with Men Who Use Inhaled Nitrites. AIDS Behav 2018 Nov;22(11):3658-3672 [FREE Full text] [doi: 10.1007/s10461-018-2139-9] [Medline: 29744767]
39. Xu J, Zhang C, Hu Q, Chu Z, Zhang J, Li Y, et al. Recreational drug use and risks of HIV and sexually transmitted infections among Chinese men who have sex with men: Mediation through multiple sexual partnerships. BMC Infect Dis 2014 Dec 02;14:642 [FREE Full text] [doi: 10.1186/s12879-014-0642-9] [Medline: 25443542]
40. Zhao P, Tang S, Wang C, Zhang Y, Best J, Tangthanasup TM, et al. Recreational Drug Use among Chinese MSM and Transgender Individuals: Results from a National Online Cross-Sectional Study. PLoS One 2017;12(1):e0170024 [FREE Full text] [doi: 10.1371/journal.pone.0170024] [Medline: 28107391]

Abbreviations

aHR: adjusted hazard ratio
HIVST: HIV self-testing
HR: hazard ratio
PrEP: pre-exposure prophylaxis
STI: sexually transmitted infection

Edited by R Kukafka; submitted 24.09.20; peer-reviewed by A Botero Carvajal, D Su; comments to author 08.11.20; revised version received 03.01.21; accepted 16.03.21; published 27.08.21.

Please cite as:
Zhang J, Tucker J, Tang W, Wang H, Chu Z, Hu Q, Huang X, Chen Y, Wang H, He X, Li Y, Zhang L, Hu Z, Bao R, Li S, Li H, Ding H, Jiang Y, Geng W, Xu J, Shang H
Internet-Based HIV Self-Testing Among Men Who Have Sex With Men Through Pre-exposure Prophylaxis: 3-Month Prospective Cohort Analysis From China
J Med Internet Res 2021;23(8):e23978
URL: https://www.jmir.org/2021/8/e23978
doi:10.2196/23978
PMID:34448709

©Jing Zhang, Joseph Tucker, Weiming Tang, Hongyi Wang, Zhenxing Chu, Qinghai Hu, Xiaojie Huang, Yaokai Chen, Hui Wang, Xiaoqing He, Yao Li, Lukun Zhang, Zhili Hu, Rantong Bao, Shangcao Li, Hang Li, Haibo Ding, Yongjun Jiang, Wenqing Geng, Junjie Xu, Hong Shang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

The Effectiveness of a Web-Based Self-Help Program to Reduce Alcohol Use Among Adults With Drinking Patterns Considered Harmful, Hazardous, or Suggestive of Dependence in Four Low- and Middle-Income Countries: Randomized Controlled Trial

Michael P Schaub[1], PhD; Marcela Tiburcio[2], PhD; Nora Martínez-Vélez[2], PhD; Atul Ambekar[3], MD; Roshan Bhad[3], MD; Andreas Wenger[1], MSc; Christian Baumgartner[1], MSc; Dzianis Padruchny[4], MSc; Sergey Osipchik[4], MD; Vladimir Poznyak[5], MD; Dag Rekve[5], MA; Fabricio Landi Moraes[6], BTech; André Luiz Monezi Andrade[7], PhD; Maria Lucia Oliveira Souza-Formigoni[6], PhD; WHO E-Health Project On Alcohol And Health Investigators Group[5]

[1]Swiss Research Institute for Public Health and Addiction, University of Zurich, Zurich, Switzerland
[2]Department of Social Sciences in Health, Ramón de la Fuente Muñiz, National Institute of Psychiatry, Mexico City, Mexico
[3]National Drug Dependence Treatment Centre, All India Institute of Medical Sciences, New Delhi, India
[4]Republican Research and Practice Center for Mental Health, Minsk, Belarus
[5]Department of Mental Health and Substance Abuse, World Health Organization, Geneva, Switzerland
[6]Departamento de Psicobiologia, Escola Paulista de Medicina, Universidade Federal de São Paulo, São Paulo, Brazil
[7]Centre for Life Sciences, Department of Psychology, Pontifica Universidade Católica de Campinas, Campinas, Brazil

Corresponding Author:
Michael P Schaub, PhD
Swiss Research Institute for Public Health and Addiction
University of Zurich
Konradstrasse 32
Zurich, 8005
Switzerland
Phone: 41 44 448 22 60
Fax: 41 44 448 22 70
Email: michael.schaub@isgf.uzh.ch

Abstract

Background: Given the scarcity of alcohol prevention and use disorder treatments in many low- and middle-income countries (LMICs), the World Health Organization has launched an eHealth portal that includes the web-based self-help program “Alcohol e-Health.”

Objective: We aimed to test the effectiveness of the Alcohol e-Health program in a randomized controlled trial.

Methods: This was a two-arm, individually randomized, and controlled trial across four LMICs comparing the self-help program and a psychoeducation and internet access as usual waiting list. Participants were broadly recruited from community samples in Belarus, Brazil, India, and Mexico from January 2016 through January 2019. The primary outcome measure was change in the Alcohol Use Disorders Identification Test (AUDIT) score with a time frame of 6 months between baseline and follow-up. Secondary outcomes included self-reported numbers of standard drinks over the previous week and cessation of harmful or hazardous drinking (AUDIT score <8).

Results: For this study, we recruited 1400 predominantly male (n=982, 70.1%) at least harmful or hazardous alcohol drinkers. The mean age was 37.6 years (SD 10.5). The participants were recruited from Brazil (n=587), Mexico (n=509), India (n=212), and Belarus (n=92). Overall, complete case analysis identified higher AUDIT changes in the intervention group (B=−4.18, 95% CI −5.42 to −2.93, P<.001, d=0.56) that were mirrored by changes in weekly standard drinks (B=−9.34, 95% CI −15.90 to −2.77, P=.005, d=0.28) and cessation rates for harmful or hazardous drinking (χ^2_1=14.56, N=561, P<.001). The supplementary intention-to-treat analyses largely confirmed these initial results.

Conclusions: The expansion of the Alcohol e-Health program to other LMICs with underdeveloped alcohol prevention and treatment systems for alcohol use disorders should be considered after successful replication of the present results.

Trial Registration: ISRCTN ISRCTN14037475; https://www.isrctn.com/ISRCTN14037475

International Registered Report Identifier (IRRID): RR2-10.1111/add.14034

(J Med Internet Res 2021;23(8):e21686) doi:10.2196/21686

KEYWORDS

alcohol; internet; public health; self-help; World Health Organization

Introduction

There is increasing interest in differences in patterns of alcohol use between low- and middle-income countries (LMICs) and high-income countries (HICs). Although alcohol use might be less common in LMICs, its risks are often more pronounced as mainly adults of low socioeconomic status [1] in these countries engage more in drinking of spirits to intoxication, a pattern very harmful to health [2]. Moreover, while alcohol is consumed by most middle-aged men and women in HICs, most users (83%) in LMICs are younger men [2]. Typical countries that follow these patterns, according to the World Health Organization (WHO) global status report on alcohol and health 2018 [3], are Belarus, Brazil, Mexico, and India (Table 1).

Table 1. Alcohol use and use disorder indicators for the four low- and middle-income countries of interest in 2016, according to the World Health Organization global status report on alcohol and health 2018 [3].

Country	Consumption of spirits as a proportion (%) of total alcoholic beverages recorded per capita in the last 12 months[a]	Heavy episodic drinking in the last 30 days (%)[a]		Alcohol use disorder prevalence (%)[b]	
		Male	Female	Male	Female
Belarus	49.0	40.5	12.2	33.9	6.2
Brazil	34.0	32.6	6.9	6.9	1.6
Mexico	20.0	30.6	6.1	4.3	0.4
India	92.0	28.4	5.4	9.1	0.5

[a] Age 15+ years.

[b] 12-month prevalence estimates including alcohol dependence and harmful use of alcohol (age 15+ years).

There are insufficient studies summarizing treatment coverage and treatment demand for alcohol use disorders in LMICs. In a systematic review, only five of the 84 low-income countries or LMICs (applying the World Bank classification) provided sufficient epidemiological data regarding harmful alcohol use [1]. According to the World Mental Health Survey, failure and delays in treatment for substance use disorders were significantly greater in LMICs compared to HICs [4]. Treatment in LMICs mainly focuses on tertiary care services for alcohol dependence with often poor outcomes [5]. Research among brief alcohol interventions in LMICs suggests that brief interventions, mainly motivational interviewing (MI) following positive screening with a standardized instrument, can help reduce self-reported hazardous or harmful alcohol use in primary care populations [6]. However, implementing brief alcohol interventions in practice and scaling them up is challenging, given a number of factors like fear of stigmatization for the patient, time pressure on the practitioner, and lack of funding for adequate training at the system level [7], which makes internet-based interventions so attractive [8].

Internet interventions that target harmful or hazardous alcohol use have been developed for HICs. In meta-analyses, the largest effect sizes were reported for studies integrating different treatment principles such as cognitive behavioral therapy (CBT), principles of self-control (PSC), and personalized normative feedback (PNF) [9,10]. These studies were superior to interventions employing elements of PNF as stand-alone interventions [9]. As expected, the first cost-effectiveness studies on internet programs for alcohol abuse were very positive. A study in a treatment center demonstrated the superiority of internet-based therapy over internet-based self-help regarding value for money when considering quality-adjusted life years gained [11]. Access to an alcohol reduction website via a brochure resulted in less costs with no worse outcome when compared to a standard face-to-face brief intervention in primary care settings [12]. However, there is a paucity of research among internet-based preventative and treatment self-help programs targeting hazardous or harmful alcohol use and alcohol use disorders [9] and mental health in general [13] in LMICs.

On December 6, 2012, the WHO launched the WHO eHealth portal for alcohol and alcohol-related consequences on health, as part of activities to reduce hazardous and harmful drinking in populations, following the objectives of the WHO's global strategy to reduce the harmful use of alcohol [14]. The portal includes the web-based self-help program called "Alcohol e-Health," an evidence-based intervention initially developed in the Netherlands [15] as a means to reduce harmful or hazardous alcohol use and use suggestive of dependence. This program has been completely revised and implemented by the WHO Department of Mental Health and Substance Use, with institutes and organizations in Belarus, Brazil, India, and Mexico. The revised intervention's effectiveness was tested in a randomized controlled trial (RCT) across the four involved countries [16]. The study's primary hypothesis was that Alcohol

XSL•FO RenderX

e-Health program participants would exhibit greater reductions in their Alcohol Use Disorders Identification Test (AUDIT) score (primary outcome [17]) between baseline and a 6-month follow-up than control subjects allocated to psychoeducation and access to the internet as usual.

Methods

Design

This study compared the Alcohol e-Health program, based on CBT, MI, and PSC, with an assessment, psychoeducation, and access to the internet as usual control group for reducing alcohol use disorder (AUDIT score) in an individually randomized controlled four LMIC trial.

The trial was executed in compliance with the Helsinki Declaration, and approved by the WHO Ethics Review Committee in October 2015 (RPC756) and four country-specific ethics committees. The study has been registered at Current Controlled Trials (registration number: ISRCTN14037475), and the detailed study protocol was published on October 26, 2017 [16].

Participants and the Inclusion and Exclusion Criteria

Participants were broadly recruited in community samples from Belarus, Brazil, India, and Mexico via information flyers and newspapers, magazines, radio, social media, websites, and informational events related to alcohol and health from January 2016 through January 2019. This broad recruitment strategy allowed for different recruitment conditions in the participating countries. However, there were considerable delays initiating recruitment in India due to official study approval procedures and changes in the study team. Moreover, continuous recruitment in Belarus proceeded very slowly. The study inclusion and exclusion criteria, and the rationale behind them are summarized in Table 2.

Table 2. Overview of the study inclusion and exclusion criteria, and the rationale behind them.

Criteria	Rationale
Inclusion criteria	
Age between 18 and 75 years	To ensure a minimal age of participation
A resident of one of the participating pilot countries	To be covered by local ethics board approval
At least weekly internet access	To ensure at least minimal program access
A screening AUDIT[a] score ≥8	To include adults with potentially hazardous or harmful alcohol consumption, and those whose drinking habits are suggestive of dependence
Exclusion criteria	
Current substance abuse treatment	To avoid confounding treatment effects
Use of opioids, inhalants, cocaine/crack or amphetamine/amphetamine-like stimulants, sedatives over the past month, or cannabis/synthetic cannabinoids for more than 4 days over the past month	To prevent confounding effects with other frequently used mind-altering drugs

[a] AUDIT: Alcohol Use Disorders Identification Test.

Intervention and Comparator

Subjects in the active study arm participated in the Alcohol e-Health program, while controls were assigned to a “waiting list,” where they were offered general information on alcohol and its effects on health, and access to the internet as usual. Program access was granted 6 months later.

The Alcohol e-Health program is an accessible interactive self-help tool for people seeking to reduce or discontinue their use of alcohol (Multimedia Appendix 1). Participants can register and use the program in their own time and free of charge. Alcohol e-Health provides support to encourage individuals to think about their drinking habits, decide whether to change their drinking behaviors, set goals regarding their drinking, take action toward reducing or stopping their drinking, measure their progress, and deal with relapses.

The core element of the program was a comprehensive diary, where participants could record every consumption occasion in terms of when, where, what, and how much they drank; with whom and how they feel about it; and other comments. Consumption of alcoholic beverages was measured in standard drinks as per the WHO Audit definition (10 g of alcohol), and a calculator assisted participants to convert their drinks into standard drinks. The drinking diary was filled out daily by dragging and dropping icons representing country-specific common drinks. Furthermore, participants could set goals for maximum number of standard drinks each day. Diary data were used for tailored weekly feedback with respect to meeting these drinking goals. Feedback was generated automatically based on individual drinking goals and country standards for low-risk drinking.

After having set baseline benchmarks regarding drinking, participants could explore the advantages and disadvantages of drinking and subsequently analyze their motivation to change. Furthermore, they were encouraged to identify risky situations with suggestions to deal with them and motivational strategies to help them maintain higher levels of resistance under such circumstances (development of alternative action plans). A relapse module allowed participants to analyze these situations (where and when, with whom, thoughts, personal feelings and thoughts, and consequences) and to cope with potential relapse (development of alternative helpful thoughts and actions in these situations). Graphical summaries of entries supported users to quickly identify core elements of risky situations. In the persistence module, users explored how they could resist social pressures to drink excessively. To address technical problems that occurred during their participation in the program, participants were permitted to contact, by email in their native language, a technician qualified to deliver technical assistance. A detailed program description is provided in the study protocol paper (Multimedia Appendix 2) [16].

Conversely, those within the psychoeducation and internet access as usual control group were told that they would be provided access to the program in 6 months, and referred to a web page containing information about the various types of alcoholic beverages, standard drink definitions, effects of alcohol on the mind and body, social effects of drinking alcohol, risk factors for alcohol dependence, women and alcohol, and adolescent alcohol use.

Throughout the 6-week program, all participants were encouraged to see a health professional if they experienced acute alcohol withdrawal or other severe physical or mental symptoms, and were afforded access to a country-specific medical advisory and emergency list.

Measurement Instruments

The main outcome was change in the adjusted AUDIT [17] score between baseline and a 6-month follow-up. Corresponding AUDIT versions were provided in English, Portuguese, Russian, and Spanish. Since the follow-up period was limited to 6 months, the AUDIT was assessed for the last 6 months instead of the last 12 for items 9 and 10, both at baseline and follow-up [17].

Secondary outcomes were as follows (Table 3): (1) falling below the cutoff of hazardous or harmful alcohol use (AUDIT score <8); (2) weekly number of standard drinks (based on a single question with seven answering fields asking about alcohol use, in standard drinks, on each day of a typical week); and (3) program satisfaction, rated using the 8-item Client Satisfaction Questionnaire (CSQ-8) [18] (assessed 6 weeks after baseline). At the 6-month follow-up, participants were asked to grade any negative effects they had experienced, as per the report by Rozental et al [19], and if they had received any external help.

Table 3. Overview of the study measurements.

Assessments/instruments	Baseline[a]	Week 6 follow-up	Month 6 follow-up[a]
Sociodemographics	Yes	No	No
AUDIT[b] score	Yes	No	Yes
Weekly number of standard drinks[c]	Yes	No	Yes
8-item Client Satisfaction Questionnaire	No	Yes	No
Adverse effects	No	No	Yes

[a] Where “Yes” is indicated both at baseline and the 6-month follow-up, the outcome of interest is the change between baseline and the 6-month follow-up.

[b] AUDIT: Alcohol Use Disorders Identification Test.

[c] Based on a single question with seven answering fields asking about alcohol use, in standard drinks, on each day of a typical week.

Sample Size

The initial sample size estimate (Cohen d=0.40 [14], 95% confidence [α=.05], and 95% power [$1-\beta$=0.95]) for analysis of covariance with one covariate (country) was 708 overall, when controlling for cluster effects [16]. However, as we met considerable follow-up problems in Brazil (see Multimedia Appendix 3 for country-specific trial flows) and as study implementation in India was delayed for technical reasons, we started to recruit only in Brazil and Mexico, with over-recruitment by 50% for the increased missing follow-up data (achieved N with 52% over-recruitment for Brazil and Mexico, including some data from Belarus [n=92], equaling 1188). However, recruitment started in India in January 2019 (n=212). Finally, the study team decided to include participants from India to increase validity for a LMIC. This resulted in a total N of 1400.

Study Procedures, Screening, and Consent

Once potential participants arrived on the Alcohol e-Health program home page, they were asked to complete the AUDIT and subsequently received personalized feedback, according to their individual drinking level concerning nonrisky drinking (<8 drinks/week), potentially harmful or hazardous drinking (8-19 drinks/week), and drinking suggestive of dependence (≥20 drinks/week). Those with an AUDIT score ≥8 were given details about the study, including (1) study aims and duration; (2) inclusion and exclusion criteria (Table 2); (3) two different study conditions; (4) potential risks of participation and safety agreements; (5) information that Alcohol e-Health cannot replace face-to-face interventions; (6) information on how participation was entirely voluntary and on their right to withdraw from the study at any time without consequences; and (7) information that the study had been approved by the WHO Research Ethics Committee and the four country-specific ethics committees. Informed consent was assumed when they selected all the necessary fields on the online informed consent form and clicked the submission icon. After having consented, participants filled out baseline questionnaires and were randomized, by a computer, to either the Alcohol e-Health program or control group, in a 1:1 ratio in each country [16]. This nonblinded assignment was registered automatically in the background database.

Follow-Up Procedure and Compensation

The 6-month follow-up assessment was performed using a step-wise procedure including electronic follow-up and reminders, as well as telephone interviews by study collaborators in their own language. Completion of all follow-up assessments qualified subjects for participation in a raffle to win a tablet or corresponding donation to a charitable organization in each country, except Brazil owing to Ethics Committee restrictions.

Statistical Analysis

For the baseline analysis between centers, we used chi-square analysis, analysis of variance (ANOVA), or the Kruskal-Wallis test, depending on the type of variable. The main analysis was based on complete case analyses (CCAs) and supplemented with intention-to-treat (ITT) analyses. Multiple linear models using regression analysis predicting the variable change score (baseline − follow-up) were used with baseline variables and the treatment condition as predictors. The missing data for ITT analyses were supplemented with multiple imputations and the R package MICE [20]. MICE specifies a multivariate distribution for missing data and draws imputations from their conditional distributions employing Markov chain Monte Carlo techniques. The missing data were assumed to be at random, and a total of 20 data sets were imputed in the supplemental ITT analyses. The effect sizes of the program were calculated using pooled results of linear model analysis. Sociodemographic (group, sex, age, country, and treatment center), and primary (AUDIT) and secondary (standard drinks in the last 7 days) outcome variables were included in the ITT imputation model. Treatment center was not used as a covariable in the overall analysis as the explained variance was less than 4%. Secondary analysis was based on complete cases, as the imputation of satisfaction scores and adverse effects was deemed unreasonable with the exception of the analysis of the number of standard drinks consumed in the last 7 days, which was additionally performed with imputed data.

Dealing With Invalid Data/Outlier Data

Outlier data were removed to increase data validity, based on the number of standard drinks consumed in the last 7 days, as there were no upper limits for self-reported data. Values above three standard deviations from the mean were considered unlikely and potentially wrong; 18 data points were set to null based on this criterion. Multimedia Appendix 4 shows a sensitivity analysis with setting of the 18 data points to the maximum remaining value in the data set.

Results

Demographics and Baseline Characteristics

Of the 1422 persons who signed up, 22 failed to complete the baseline survey. Thus, a total of 1400 participants (418 female, 29.9%) were randomly allocated to two study arms (Figure 1). Table 4 summarizes the demographic characteristics and baseline screening data of and statistical comparisons between the four study centers. The average age of participants was 37.6 years (SD 10.5). Participants consumed an average of 43.7 (SD 41.4) standard drinks per week and had an average of 2.6 (SD 2.0) alcohol-free days per week at baseline. The average AUDIT score at baseline was 23.0 (SD 7.7), with significant differences between the four centers for all variables.

Figure 1. CONSORT-EHEALTH flowchart: overview of participant flow. CCA: complete case analysis; ITT: intention to treat.

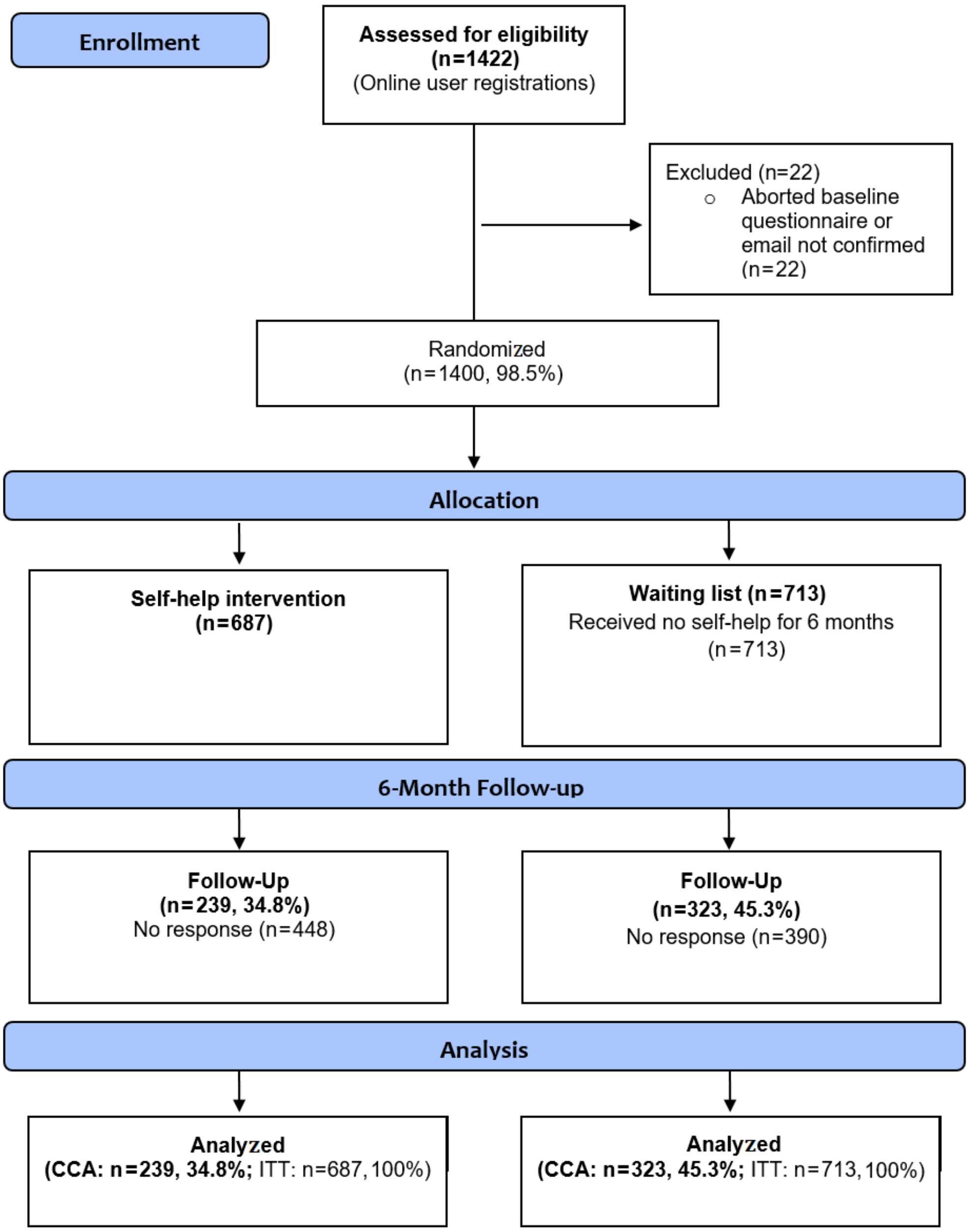

Enrollment
Assessed for eligibility (n=1422) (Online user registrations)
Excluded (n=22)
○ Aborted baseline questionnaire or email not confirmed (n=22)
Randomized (n=1400, 98.5%)
Allocation
Self-help intervention (n=687)
Waiting list (n=713) Received no self-help for 6 months (n=713)
6-Month Follow-up
Follow-Up (n=239, 34.8%) No response (n=448)
Follow-Up (n=323, 45.3%) No response (n=390)
Analysis
Analyzed (CCA: n=239, 34.8%; ITT: n=687, 100%)
Analyzed (CCA: n=323, 45.3%; ITT: n=713, 100%)

Table 4. Demographic and baseline characteristics of all centers and study arms.

Characteristic	Brazil (N=587)		Mexico (N=509)		India (N=212)		Belarus (N=92)		Total	Statistical value[a]	*P* value
	IG[b] (n=290)	CG[c] (n=297)	IG (n=256)	CG (n=253)	IG (n=95)	CG (n=117)	IG (n=46)	CG (n=46)			
Gender, n (%)										χ^2 (3)=68.9, N=1400	<.001
Female	99 (34.1%)	125 (42.1%)	57 (22.3%)	82 (32.4%)	10 (10.5%)	9 (7.7%)	16 (34.8%)	20 (43.5%)	418 (29.9%)		
Male	191 (65.9%)	172 (57.9%)	199 (77.7%)	171 (67.6%)	85 (89.5%)	108 (92.3%)	30 (65.2%)	26 (56.5%)	982 (70.1%)		
Age (years), mean (SD)	37.6 (10.6)	36.6 (10.2)	36.6 (11.0)	36.4 (10.6)	38.6 (9.5)	40.3 (8.9)	43.1 (9.4)	41.0 (10.7)	37.6 (10.5)	*F* (3)=10.4, N=1400	<.001
Alcohol Use Disorders Identification Test (AUDIT) score, mean (SD)	22.3 (6.8)	22.2 (6.5)	22.6 (6.3)	22.3 (6.8)	30.2 (7.2)	30.2 (8.6)	13.1 (4.1)	14.4 (5.7)	23.0 (7.7)	*F* (3)=142.2, N=1400	<.001
Standard drinks[d], mean (SD)	44.6 (29.6)	42.4 (28.6)	28.3 (18.4)	30.4 (19.6)	93.0 (71.2)	90.5 (69.1)	14.5 (14.4)	13.7 (12.4)	43.7 (41.4)	*F* (3)=178.4, N=1382	<.001
Drinking-free days[d], mean (SD)	2.8 (2.0)	2.2 (2.2)	3.2 (2.0)	2.4 (2.0)	0.4 (1.3)	0.2 (0.4)	3.0 (1.3)	2 (0.8)	2.6 (2.0)	*F* (3)=51.0, N=703	<.001

[a] Chi-square test, analysis of variance, or Kruskal-Wallis test.

[b] IG: intervention group.

[c] CG: control group.

[d] Last 7 days.

Study Attrition and Dropout Analysis

In total, 562 (40.1%) participants completed follow-up. Dropout analysis showed a significant difference by country of enrolment (χ^2_3=36.88, N=1400, *P*<.001), assigned study condition (χ^2_1=16.11, N=1400, *P*<.001), baseline AUDIT score (F_1=9.69, N=1400, *P*=.002), and baseline drink-free days over the last 7 days (F_1=6.94, N=701, *P*=.009). The follow-up rate ranged from 27.6% (n=80) in the intervention group in Brazil to 56.2% (n=26) in the control group in Belarus, and in total, 34.9% (n=239) of the intervention group could be followed compared to 45.3% (n=323) of the control group. The average AUDIT score and number of drinking-free days were 22.18 (SD 7.92) and 2.35 (SD 2.07), respectively, for participants who could be followed up, and 23.48 (SD 7.49) and 2.77 (SD 2.02), respectively, for dropouts. Dropouts were not different in gender (χ^2_1=0.13, N=1400, *P*=.72), age (F_1=2.50, N=1398, *P*=.11), baseline standard drinks over the preceding 7 days (F_1=3.32, N=1380, *P*=.069), or adherence ($t_{417.5}$=1.76, *P*=.08). Detailed follow-up rates are listed in Multimedia Appendix 3.

Main Effects

Table 5 shows the detailed outcome analysis of changes in the AUDIT score (main outcome) and changes in the number of standard drinks in the last 7 days across all countries. Regression analysis based on the complete case data showed a significant decrease in the AUDIT score in the intervention group (mean 7.4, SD 7.8) compared with that in the control group (mean 3.2, SD 7.1) (*B*=−4.18, 95% CI −5.42 to −2.93, *P*<.001, *d*=0.56). Table 6 shows detailed values between baseline and follow-up from complete cases, and Multimedia Appendix 5 and Multimedia Appendix 6 show values from supplemental imputed data that confirmed the CCA results.

Table 5. Regression analysis results.

Variable	Intervention versus control after 6 months (complete cases) (N=562)		
	B [a]	95% CI	*P* value
AUDIT[b]	−4.18	−5.42 to −2.93	<.001
Standard drinks[c]	−9.34	−15.90 to −2.77	.005

[a] Baseline data and condition as predictors for group effect.

[b] AUDIT: Alcohol Use Disorders Identification Test.

[c] Last 7 days.

Table 6. Values between baseline and follow-up from complete cases.

Variable	Control baseline (n=713), mean (SD)	Intervention baseline (n=687), mean (SD)	Control followed up[a] (n=325), mean (SD)	Intervention followed up[a] (n=239), mean (SD)	d [b]	95% CI
AUDIT[c]	23.05 (7.88)	22.86 (7.50)	18.71 (9.28)	15.15 (9.06)	0.56	0.38-0.72
Standard drinks[d]	44.21 (41.70)	43.23 (41.13)	23.73 (26.32)	12.46 (16.31)	0.28	0.08-0.46
CSQ-8[e]	N/A[f]	N/A	18.92 (4.65)	21.56 (4.11)	0.60	0.40-0.79

[a] 6 months after baseline (complete cases).

[b] Effect size Cohen *d* based on differences between the intervention and control groups.

[c] AUDIT: Alcohol Use Disorders Identification Test.

[d] Last 7 days.

[e] CSQ-8: 8-item Client Satisfaction Questionnaire.

[f] N/A: not applicable.

Secondary Outcomes

Standard Drinks

There was a significantly higher decrease in standard drinks in the intervention group (mean 24.7, SD 39.6) compared with that in the control group (mean 15.4, SD 28.4) (B=−9.34, 95% CI −15.90 to −2.77, P=.005, d=0.28). Results were confirmed in the supplemental ITT analysis (see Multimedia Appendix 5 and Multimedia Appendix 6).

Harmful or Hazardous Alcohol Use

A total of 102 (18.2%) participants had a total AUDIT score below 8 after 6 months and could therefore be classified as no more harmful or hazardous alcohol use. There was a significant difference in the assigned group, with 57 (25.6%) in the intervention group versus 41 (12.7%) in the control group achieving a score below 8 after 6 months (χ^2_1=14.56, N=561, P<.001).

CSQ-8

Participants in the intervention group were significantly more satisfied with their study participation (mean CSQ-8 score 21.56, SD 4.1) compared with those allocated to the control group (mean score 18.92, SD 4.7) ($t_{415.39}$=6.18, P<.001).

Adherence

Of the 687 participants in the intervention group, 258 (37.6%) had at least one diary entry, 159 (23.1%) completed the tools to maintain their targets for alcohol consumption and to resist social pressure, and 41 (6.0%) completed the relapse tool.

Adverse Effects

A total of 188 participants completed the Rozental adverse side effects questionnaire [17]. Of these, 136 (71.9%) answered that they had not experienced any negative effects during the study, while 33 (17.5%) claimed that an adverse effect had affected them "somewhat negatively," 7 (3.7%) claimed "quite negatively," and 13 (6.9%) claimed "to a great extent." There was no significant intergroup difference ($t_{151.93}$=1.35, P=.18).

Discussion

Principal Findings

To our knowledge, this is the first RCT to investigate the effectiveness of an international web-based self-help program for at least harmful or hazardous drinkers in LMICs. The observed changes in AUDIT scores (main outcome) were mirrored by all of the secondary outcomes in the CCA and mostly confirmed in the supplemental ITT analyses. Adverse effects credited to the intervention were few. Thus, expanding this program to other LMICs with underdeveloped alcohol prevention and alcohol use disorder treatment systems should be considered after successful replication of these initial findings. The achieved effect strength for ITT changes in the weekly number of standard drinks (d=0.27) in this study is only slightly smaller than that reported for meta-analyses from HICs (d=0.40) [9] and evaluation of the initial Dutch program (d=0.40) [15] from which development of the Alcohol e-Help program started [16]. However, effects on the main outcome, change in the AUDIT score, were even higher (CCA: d=0.55; ITT: d=0.51). Presumably, in this study, the standardized AUDIT scores [17] with validated translations were more appropriate as an outcome measure than the number of weekly standard drinks, despite country-specific adaptation of the standard drink examples.

The achieved effects were mainly grounded in the Brazilian and Mexican middle-income country data, but were also observed in data from India, the only low-income country involved. Moreover, the program attracted subjects with a high probability of alcohol dependence (mean AUDIT score above 22) in these countries. In India, the baseline AUDIT score was 30. Our Indian collaborators noted that recruitment was difficult in the general population and that some of the participants might have been reached through community clinic settings where recruitment posters were sometimes hung, which may partially explain their higher baseline scores. The only country where we failed to demonstrate program effectiveness was Belarus, the middle-income country with the highest spirit use and clearly the highest alcohol use disorder prevalence. Unfortunately, we only reached few Belarusian participants with, contrary to our expectations, a comparatively low level of alcohol use and very few with an AUDIT level suggestive of dependence. Our

Belarusian collaborators have advised us that there is no sense of harmful and hazardous alcohol use in the general Belarusian population and spirit use is culturally still very accepted. Unless a medical doctor refers someone for inpatient detoxification, many Belarussians with high levels of spirit use still seem to feel they do not have a health problem warranting changes in their alcohol drinking behaviors. Thus, a comprehensive prevention campaign to make the Belarussian drinking population aware of potential health problems and the consequences of their drinking behaviors might be a required preliminary step before implementation of the Alcohol e-Help program or similar interventions. In addition, a cultural adaptation of the program to reflect these values and attitudes toward excessive alcohol consumption could also be important. This could also apply to other former Soviet Union regions with similar drinking cultures and, thus, limits the generalizability of our study results. Generalizability is further limited to individuals with sufficient reading and writing skills in the corresponding study languages, those with at least weekly internet access, and those without frequent use of illicit drugs.

Despite the typical gender gap in heavy episodic drinking and prevalence of alcohol use disorders in LMICs compared to HICs [1,2], the Alcohol e-Help program still reached a population that was almost 30.0% female, and dropout rates were similar in the two genders. It is possible that similar to HICs, relatively more women will be reached by online interventions compared to face-to-face interventions. A detailed analysis of the outcome predictors in the near future could clarify the relevance of gender and other predictors for the program's effectiveness.

On one hand, the decision to over-recruit made it possible to include data from the delayed recruitment in India in this study. On the other hand, over-recruitment increases the chance of detecting an intervention effect. Since the effect strengths determined were analogous or, in the case of the main outcome, significantly higher than originally calculated, most of the results would probably have been significant even with the originally calculated sample size. For this reason, we have not subsequently adjusted our power calculation.

Study recruitment was set as broad as possible to encourage people from the respective general populations with at least harmful or hazardous alcohol consumption, who otherwise would not receive adequate support, to participate. In this way, we also wanted to optimize the study results' generalizability. However, this strategy probably also led to recruitment distortions between the participating countries. In Brazil, for example, a large proportion of participants were recruited through television coverage, as the Brazilian media displayed great interest in the study. Conversely, in India, the only truly successful recruitment strategies were newspaper advertisements and hanging posters, since the Indian media would have reported on the study only if financially compensated.

One difficulty that arose due to the comparatively long study preparation and study phase was that the underlying content management system (CMS) was outdated and an update during the study phase in the four country-specific portals turned out to be tricky for reasons of methodological and data loss risks. Before the Alcohol e-Help program can be adapted for other LMICs and widely implemented, its content must be transferred to the latest CMS in a time-consuming process. However, measured against the public health impact potential and increasing spread of broadband internet access in LMICs [21], this seems to be a worthwhile investment.

A major limitation of this study is the high attrition rate of nearly 60%, which was larger than for studies in HICs [9]. In addition, selective attrition was a concern with higher dropout rates in the intervention group and more severe baseline AUDIT scores reported by noncompleters. These difficulties might compromise the stability and reliability of effects. Therefore, we based our results on the CCA and added the ITT analyses only supplementally. Up to which missing level and under which conditions data can still be imputed for ITT analysis is part of ongoing discussions [22,23]. Particularly for internet-based programs in the alcohol field, this discussion is still pending. Since follow-up surveys are obviously more difficult to conduct in low-income countries, this discussion must be held as soon as possible and a consensus should be reached in this regard. The transfer of successful programs from HICs should not be hindered merely by lack of consensus in these discussions. However, since we have carried out the first transfer of a program to reduce alcohol misuse to LMICs, there is certainly room for improvement to increase the follow-up rate. Higher financial compensation would be an obvious solution. However, we deliberately decided against this because we saw a risk of an imbalance, and thus, an uncontrollable confounder, between the different countries involved. Another possibility would be to link the follow-up survey with a content booster module, in which, for example, short personalized feedback using the AUDIT strategies for the long-term success of alcohol consumption reduction is addressed. A follow-up survey would then make more sense to some participants, and they would be more likely to participate. In addition, better imputation programs, especially for data sets with a high missing proportion, are hoped to be available soon.

Other limitations of this study are as follows: (1) self-reported use of standard drinks that might particularly be susceptible to retrospective bias, due to inclusion of individuals suggestive of dependence (AUDIT score ≥21); (2) skewed country distribution of participants with an underrepresentation of individuals from Belarus and India, which made some of the ITT analyses in these two countries impossible (Multimedia Appendix 7); and (3) incorrect programing of the originally planned adherence assessments [16]. As such, no adequate sensitivity analysis could be performed.

Conclusions

Based on the results reported here, expansion of the program to other LMICs with underdeveloped alcohol prevention and treatment systems for alcohol use disorders should be considered after successful replication of the initial results.

Acknowledgments

We thank the following additional members of the World Health Organization (WHO) e-Health project on Alcohol and Health Investigators Group from Brazil: Ana Regina Noto, Denise De Micheli, Ana Paula L Carneiro, Roseli Boerngen-Lacerda, Telmo M Ronzani, Laisa M A Sartes, and Leonardo F Martins. We thank the following members from India: Rajat Ray, Sudhir Khandelwal, Rakesh Chadda, and Urvashi Singh. We thank the following members from Mexico: María Asunción Lara, Silvia Carreño García, and Morise Fernández Torres. We thank Isy Vromans who managed the project at the initial stages and managed the development of the intervention in collaboration with the Trimbos Institute. We would also like to thank Otavio J E Pereira, Gabrielle A Cunha, Marcia Omori, and Denise Silva for their technical support. This study was funded by the WHO (APW 2018/806239-0). MLOSF received additional funding from the Associação Fundo de Incentivo à Pesquisa, CNPq (315769/2018-1), CAPES, and the São Paulo Research Foundation FAPESP (2015/19472-5).

Authors' Contributions

MPS, MT, NMV, AA, RB, AW, DP, SO, FLM, ALMA, and MLOSF significantly contributed to the development of the portal and adaptation of the intervention and/or were substantially involved in subject recruitment and data collection. MPS was the overall principal investigator, wrote the first draft of the manuscript, and analyzed the data together with AW and CB. DR and VP were the project leaders at the World Health Organization and had the overall project lead. The WHO e-Health Project on Alcohol and Health Investigators Group includes Michael P Schaub, Marcela Tiburcio, Nora Martínez-Vélez, Atul Ambekar, Yatan Pal Singh Balhara, Andreas Wenger, André Luiz Monezi Andrade, Dzianis Padruchny, Sergey Osipchik, Elise Gehring, Vladimir Poznyak, Dag Rekve, Maria Lucia Oliveira Souza-Formigoni, Ekaterina Lapushinskaya, Pavel Rynkov, Dmitry Statkevich, Roseli Boerngen Lacerda, Telmo Ronzani, Laisa M A Sartes, María Asunción Lara, Morise Fernández, Anubha Dhall, Monica Mongia, Roshan Bad, and Urvashi Raj. All authors discussed the results and contributed to the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Introduction video to the Alcohol e-Health program.
[MP4 File (MP4 Video), 17160 KB - jmir_v23i8e21686_app1.mp4]

Multimedia Appendix 2
Detailed technical description of the Alcohol e-Health program.
[PDF File (Adobe PDF File), 426 KB - jmir_v23i8e21686_app2.pdf]

Multimedia Appendix 3
CONSORT-EHEALTH trial flowchart: overview of participant flow split in study centers.
[PNG File , 192 KB - jmir_v23i8e21686_app3.png]

Multimedia Appendix 4
Sensitivity analysis replacing outliers with the highest remaining value.
[PDF File (Adobe PDF File), 253 KB - jmir_v23i8e21686_app4.pdf]

Multimedia Appendix 5
Intention-to-treat regression analysis results.
[PDF File (Adobe PDF File), 249 KB - jmir_v23i8e21686_app5.pdf]

Multimedia Appendix 6
Intention-to-treat means, standard deviations, and achieved effect sizes.
[PDF File (Adobe PDF File), 250 KB - jmir_v23i8e21686_app6.pdf]

Multimedia Appendix 7
Detailed analysis for every study country.
[PDF File (Adobe PDF File), 259 KB - jmir_v23i8e21686_app7.pdf]

Multimedia Appendix 8
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 809 KB - jmir_v23i8e21686_app9.pdf]

References

1. Allen L, Williams J, Townsend N, Mikkelsen B, Roberts N, Foster C, et al. Socioeconomic status and non-communicable disease behavioural risk factors in low-income and lower-middle-income countries: a systematic review. The Lancet Global Health 2017 Mar;5(3):e277-e289 [FREE Full text] [doi: 10.1016/S2214-109X(17)30058-X] [Medline: 28193397]
2. Smyth A, Teo KK, Rangarajan S, O'Donnell M, Zhang X, Rana P, et al. Alcohol consumption and cardiovascular disease, cancer, injury, admission to hospital, and mortality: a prospective cohort study. The Lancet 2015 Nov 14;386(10007):1945-1954. [doi: 10.1016/S0140-6736(15)00235-4] [Medline: 26386538]
3. Global status report on alcohol and health. World Health Organization. 2018. URL: https://apps.who.int/iris/bitstream/handle/10665/274603/9789241565639-eng.pdf [accessed 2021-08-04]
4. Wang PS, Angermeyer M, Borges G, Bruffaerts R, Tat Chiu W, DE Girolamo G, et al. Delay and failure in treatment seeking after first onset of mental disorders in the World Health Organization's World Mental Health Survey Initiative. World Psychiatry 2007 Oct;6(3):177-185 [FREE Full text] [Medline: 18188443]
5. Benegal V, Chand PK, Obot IS. Packages of care for alcohol use disorders in low- and middle-income countries. PLoS Med 2009 Oct;6(10):e1000170 [FREE Full text] [doi: 10.1371/journal.pmed.1000170] [Medline: 19859536]
6. Joseph J, Basu D. Efficacy of Brief Interventions in Reducing Hazardous or Harmful Alcohol Use in Middle-Income Countries: Systematic Review of Randomized Controlled Trials. Alcohol Alcohol 2017 Jan;52(1):56-64. [doi: 10.1093/alcalc/agw054] [Medline: 27567270]
7. Derges J, Kidger J, Fox F, Campbell R, Kaner E, Hickman M. Alcohol screening and brief interventions for adults and young people in health and community-based settings: a qualitative systematic literature review. BMC Public Health 2017 Jun 09;17(1):562 [FREE Full text] [doi: 10.1186/s12889-017-4476-4] [Medline: 28599632]
8. Beyer F, Lynch E, Kaner E. Brief Interventions in Primary Care: an Evidence Overview of Practitioner and Digital Intervention Programmes. Curr Addict Rep 2018;5(2):265-273 [FREE Full text] [doi: 10.1007/s40429-018-0198-7] [Medline: 29963364]
9. Riper H, Hoogendoorn A, Cuijpers P, Karyotaki E, Boumparis N, Mira A, et al. Effectiveness and treatment moderators of internet interventions for adult problem drinking: An individual patient data meta-analysis of 19 randomised controlled trials. PLoS Med 2018 Dec;15(12):e1002714 [FREE Full text] [doi: 10.1371/journal.pmed.1002714] [Medline: 30562347]
10. Hadjistavropoulos HD, Mehta S, Wilhelms A, Keough MT, Sundström C. A systematic review of internet-delivered cognitive behavior therapy for alcohol misuse: study characteristics, program content and outcomes. Cogn Behav Ther 2020 Jul;49(4):327-346. [doi: 10.1080/16506073.2019.1663258] [Medline: 31599198]
11. Blankers M, Nabitz U, Smit F, Koeter MWJ, Schippers GM. Economic evaluation of internet-based interventions for harmful alcohol use alongside a pragmatic randomized controlled trial. J Med Internet Res 2012 Oct 29;14(5):e134 [FREE Full text] [doi: 10.2196/jmir.2052] [Medline: 23103771]
12. Hunter R, Wallace P, Struzzo P, Vedova RD, Scafuri F, Tersar C, et al. Randomised controlled non-inferiority trial of primary care-based facilitated access to an alcohol reduction website: cost-effectiveness analysis. BMJ Open 2017 Nov 03;7(11):e014577 [FREE Full text] [doi: 10.1136/bmjopen-2016-014577] [Medline: 29102983]
13. Arjadi R, Nauta MH, Chowdhary N, Bockting CLH. A systematic review of online interventions for mental health in low and middle income countries: a neglected field. Glob Ment Health (Camb) 2015;2:e12 [FREE Full text] [doi: 10.1017/gmh.2015.10] [Medline: 28596860]
14. Global strategy to reduce the harmful use of alcohol. World Health Organization. 2010. URL: http://www.who.int/substance_abuse/alcstratenglishfinal.pdf [accessed 2021-08-02]
15. Riper H, Kramer J, Smit F, Conijn B, Schippers G, Cuijpers P. Web-based self-help for problem drinkers: a pragmatic randomized trial. Addiction 2008 Feb;103(2):218-227. [doi: 10.1111/j.1360-0443.2007.02063.x] [Medline: 18199300]
16. Schaub MP, Tiburcio M, Martinez N, Ambekar A, Balhara YPS, Wenger A, WHO e-Health Project on AlcoholHealth Investigators Group. Alcohol e-Help: study protocol for a web-based self-help program to reduce alcohol use in adults with drinking patterns considered harmful, hazardous or suggestive of dependence in middle-income countries. Addiction 2018 Feb;113(2):346-352. [doi: 10.1111/add.14034] [Medline: 28921778]
17. Babor TF, Higgins-Biddle J, Saunders J, Monteiro M. AUDIT: The Alcohol Use Disorders Identification Test: Guidelines for Use in Primary Care. 2nd Ed. World Health Organization. 2001. URL: https://apps.who.int/iris/handle/10665/67205 [accessed 2021-08-02]
18. Attkisson CC, Zwick R. The client satisfaction questionnaire. Psychometric properties and correlations with service utilization and psychotherapy outcome. Eval Program Plann 1982;5(3):233-237. [doi: 10.1016/0149-7189(82)90074-x] [Medline: 10259963]
19. Rozental A, Boettcher J, Andersson G, Schmidt B, Carlbring P. Negative effects of internet interventions: a qualitative content analysis of patients' experiences with treatments delivered online. Cogn Behav Ther 2015;44(3):223-236. [doi: 10.1080/16506073.2015.1008033] [Medline: 25705924]
20. Buuren SV, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. J. Stat. Soft 2011;45(3):1-67. [doi: 10.18637/jss.v045.i03]
21. The State of Broadband 2018: Broadband catalyzing sustainable development. International Telecommunication Union. 2018. URL: https://www.itu.int/dms_pub/itu-s/opb/pol/S-POL-BROADBAND.19-2018-PDF-E.pdf [accessed 2021-08-02]

22. Madley-Dowd P, Hughes R, Tilling K, Heron J. The proportion of missing data should not be used to guide decisions on multiple imputation. J Clin Epidemiol 2019 Jun;110:63-73 [FREE Full text] [doi: 10.1016/j.jclinepi.2019.02.016] [Medline: 30878639]
23. Hughes RA, Heron J, Sterne JAC, Tilling K. Accounting for missing data in statistical analyses: multiple imputation is not always the answer. Int J Epidemiol 2019 Aug 01;48(4):1294-1304 [FREE Full text] [doi: 10.1093/ije/dyz032] [Medline: 30879056]

Abbreviations

AUDIT: Alcohol Use Disorders Identification Test
CBT: cognitive behavioral therapy
CCA: complete case analysis
CMS: content management system
CSQ-8: 8-item Client Satisfaction Questionnaire
HIC: high-income country
ITT: intention to treat
LMIC: low- and middle-income country
MI: motivational interviewing
PNF: personalized normative feedback
PSC: principles of self-control
RCT: randomized controlled trial
WHO: World Health Organization

Edited by R Kukafka; submitted 26.06.20; peer-reviewed by M Blankers, F Hamilton, B Suffoletto; comments to author 20.07.20; revised version received 17.08.20; accepted 13.07.21; published 27.08.21.

Please cite as:
Schaub MP, Tiburcio M, Martínez-Vélez N, Ambekar A, Bhad R, Wenger A, Baumgartner C, Padruchny D, Osipchik S, Poznyak V, Rekve D, Landi Moraes F, Monezi Andrade AL, Oliveira Souza-Formigoni ML, WHO E-Health Project On Alcohol And Health Investigators Group
The Effectiveness of a Web-Based Self-Help Program to Reduce Alcohol Use Among Adults With Drinking Patterns Considered Harmful, Hazardous, or Suggestive of Dependence in Four Low- and Middle-Income Countries: Randomized Controlled Trial
J Med Internet Res 2021;23(8):e21686
URL: https://www.jmir.org/2021/8/e21686
doi:10.2196/21686
PMID:34448710

©Michael P Schaub, Marcela Tiburcio, Nora Martínez-Vélez, Atul Ambekar, Roshan Bhad, Andreas Wenger, Christian Baumgartner, Dzianis Padruchny, Sergey Osipchik, Vladimir Poznyak, Dag Rekve, Fabricio Landi Moraes, André Luiz Monezi Andrade, Maria Lucia Oliveira Souza-Formigoni, WHO E-Health Project On Alcohol And Health Investigators Group. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO RenderX

Original Paper

Informing Patients With Esophagogastric Cancer About Treatment Outcomes by Using a Web-Based Tool and Training: Development and Evaluation Study

Loïs F van de Water[1,2*], MSc; Héctor G van den Boorn[1*], MSc; Florian Hoxha[1], BSc; Inge Henselmans[2], PhD; Mart M Calff[2], PhD; Mirjam A G Sprangers[2], PhD; Ameen Abu-Hanna[3], PhD; Ellen M A Smets[2], PhD; Hanneke W M van Laarhoven[1], MD, PhD

[1]Department of Medical Oncology, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, Netherlands
[2]Department of Medical Psychology, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, Netherlands
[3]Department of Medical Informatics, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, Netherlands
[*]these authors contributed equally

Corresponding Author:
Hanneke W M van Laarhoven, MD, PhD
Department of Medical Oncology
Amsterdam University Medical Centers
University of Amsterdam
Meibergdreef 9
Amsterdam,
Netherlands
Phone: 31 205665955
Fax: 31 206919743
Email: h.vanlaarhoven@amsterdamumc.nl

Abstract

Background: Due to the increasing use of shared decision-making, patients with esophagogastric cancer play an increasingly important role in the decision-making process. To be able to make well-informed decisions, patients need to be adequately informed about treatment options and their outcomes, namely survival, side effects or complications, and health-related quality of life. Web-based tools and training programs can aid physicians in this complex task. However, to date, none of these instruments are available for use in informing patients with esophagogastric cancer about treatment outcomes.

Objective: This study aims to develop and evaluate the feasibility of using a web-based prediction tool and supporting communication skills training to improve how physicians inform patients with esophagogastric cancer about treatment outcomes. By improving the provision of treatment outcome information, we aim to stimulate the use of information that is evidence-based, precise, and personalized to patient and tumor characteristics and is communicated in a way that is tailored to individual information needs.

Methods: We designed a web-based, physician-assisted prediction tool—Source—to be used during consultations by using an iterative, user-centered approach. The accompanying communication skills training was developed based on specific learning objectives, literature, and expert opinions. The Source tool was tested in several rounds—a face-to-face focus group with 6 patients and survivors, semistructured interviews with 5 patients, think-aloud sessions with 3 medical oncologists, and interviews with 6 field experts. In a final pilot study, the Source tool and training were tested as a combined intervention by 5 medical oncology fellows and 3 esophagogastric outpatients.

Results: The Source tool contains personalized prediction models and data from meta-analyses regarding survival, treatment side effects and complications, and health-related quality of life. The treatment outcomes were visualized in a patient-friendly manner by using pictographs and bar and line graphs. The communication skills training consisted of blended learning for clinicians comprising e-learning and 2 face-to-face sessions. Adjustments to improve both training and the Source tool were made according to feedback from all testing rounds.

Conclusions: The Source tool and training could play an important role in informing patients with esophagogastric cancer about treatment outcomes in an evidence-based, precise, personalized, and tailored manner. The preliminary evaluation results are promising and provide valuable input for the further development and testing of both elements. However, the remaining uncertainty about treatment outcomes in patients and established habits in doctors, in addition to the varying trust in the prediction models, might influence the effectiveness of the tool and training in daily practice. We are currently conducting a multicenter clinical trial to investigate the impact that the combined tool and training have on the provision of information in the context of treatment decision-making.

(J Med Internet Res 2021;23(8):e27824) doi:10.2196/27824

KEYWORDS

prediction tool; communication skills training; shared decision-making; risk communication; treatment outcomes; esophageal cancer; gastric cancer; patient-physician communication

Introduction

Background

Esophageal and gastric cancers rank eighth and fifth, respectively, in incidence worldwide [1]. The mortality rate is high and, even in the curative setting, the 5-year survival rates do not exceed 50% [2,3]. Over the years, several treatment regimens have come into use, resulting in an array of treatments varying in their effectiveness regarding survival, health-related quality of life (HRQoL), and side effects and complications. For example, localized esophageal cancer can be treated with resection, with or without neoadjuvant chemotherapy or chemoradiotherapy, or with definitive chemoradiation, and localized gastric cancer can be treated with resection with or without adjuvant or neoadjuvant chemotherapy [4-6]. Various options exist for metastasized cancers, with chemotherapy yielding the best survival rates. However, palliative radiotherapy and best supportive care may also be valuable options for specific groups of patients [4,5,7-10].

Oftentimes, the choice between treatment options is based on preferences; the personal weighing of the pros and cons of the options plays a decisive role in the final decision made, and therefore, shared decision-making is needed [11,12]. For shared decision-making to be effective, patients need to be well-informed and thus be offered evidence-based and precise information on treatment outcomes. Evidence-based information refers to the best available, most accurate, and up-to-date evidence. Precise information is concrete, clear, and substantially detailed, such as “In 5 years, 45 out of 100 patients like you that are given this treatment will still be alive.” However, treatment outcomes can differ according to specific patient characteristics (such as age and performance status) and tumor characteristics (such as tumor–node–metastasis [TNM] staging and the number of metastases) [13,14]. Thus, physicians face the challenge of having to inform patients on treatment-related outcomes in a manner that is not only evidence-based and precise but also personalized to the individual patient.

Physicians may face many other challenges when informing patients with cancer on treatment and related outcomes. A vast amount of information on the possible treatment options, including their procedures and associated risks and benefits, must be communicated within the time restrictions of a consultation [15]. Moreover, this information, including schedules, numbers, and probabilities, is often complex and therefore difficult for patients to process [16,17]. Patients’ emotions can complicate information processing even further, especially as esophagogastric cancer is a life-threatening disease [18,19]. Physicians consider dealing with these emotions as a difficult-to-acquire skill [20]. They often worry that their information might even increase a patient’s anxiety or take away a patient’s hope [21-25]. Therefore, physicians may have conflicting opinions and doubts about how to provide precise and numerical information regarding treatment risks and benefits.

Furthermore, tailoring the type and amount of information to the individual patient’s information needs, interests, and concerns (eg, one patient wants to be informed using exact percentages, whereas another would rather get a general description) has also been shown to be a difficult skill for physicians [26]. These challenges impede the ability to meet the information needs of patients with cancer [27-29]. Physicians rarely use clinical outcome data to systematically inform patients, given a certain treatment, on their chances of survival, the most likely side effects, and the consequences on their quality of life [30,31]. However, it has been established that many patients want to receive more information on their treatment-related outcomes and want this information to be more precise [32-36].

Several tools have been developed to aid physicians in this task by using prediction models to generate clarifying visualizations of personalized outcome data, such as the *Predict* and *Adjuvant Online* tools for breast cancer [37,38]. To achieve personalized prediction, these models use multiple characteristics of the patient and the disease to create bar plots and Kaplan-Meier curves displaying survival data. However, to date, no web-based prediction tool exists for use in clinical consultations targeted at patients with esophageal and gastric cancer [39]. Moreover, the probabilities of side effects and HRQoL related to the treatment options are not addressed in the current tools, although patients express information needs related to these outcomes [32,40]. Furthermore, several training programs are available to improve the communication skills of cancer care providers [41,42]. However, these often do not specifically address how to inform patients about treatment options and their particular outcomes, for instance, by using a prediction tool. Combining a prediction tool with communication skills training to address knowledge, attitudes, and skills might increase the usage and adoption of the new tool in clinical practice, improve the overall communication of outcome information by physicians, and stimulate shared decision-making.

Objectives

Therefore, we aim to develop a web-based prediction tool and supporting communication skills training to improve how physicians inform patients with esophagogastric cancer on treatment outcomes, namely, survival, side effects or complications, and HRQoL. To improve the provision of treatment outcome information, we aim for information that is evidence-based, precise, and personalized to the patient and tumor characteristics, and that is communicated in a way tailored to the individual information needs.

Introducing a change in physician-patient communication by adding a new instrument might initially result in resistance from users, as suggested by behavior change theories [43]. For example, physicians might be reluctant to use the tool because it does not fit into their consultation routine or because they might lack trust in the prediction models. Therefore, our secondary aim is to evaluate the feasibility of the tool and training in practice by consulting physicians, patients, survivors, and experts and to iteratively improve the tool and training.

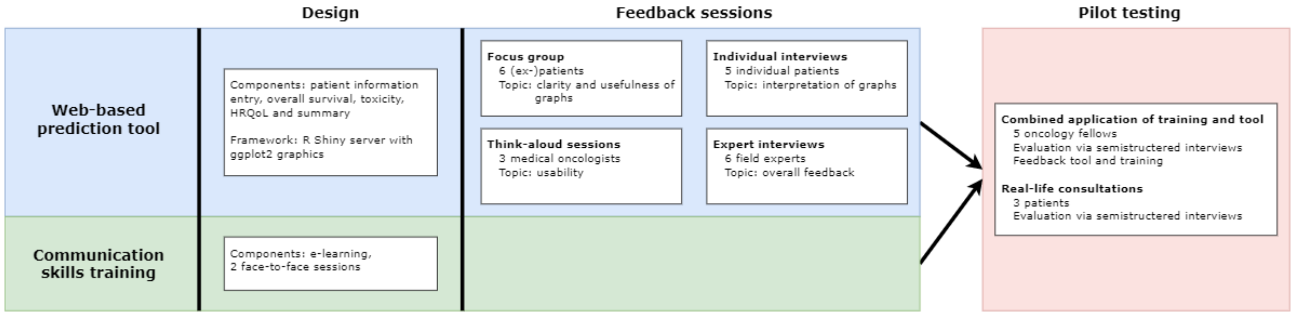
Methods

Overview

Both the tool and training were targeted at physicians in oncology who regularly conduct treatment decision-making consultations. In the development of the tool and training, we focused on patients with metastatic esophageal and gastric cancer. With regard to shared decision-making, this group is confronted with the most complex decision-making process, where personal values and preferences play a large role in deciding among multiple relevant treatment options.

The iterative development and testing of this two-part intervention occurred in several phases following the 2008 Medical Research Council framework [44]. This framework provides guidance for developing complex interventions and presents several steps and elements necessary for the successful implementation of the intervention. The framework is divided into the following four phases: (1) development, (2) piloting, (3) evaluation, and (4) implementation. This study describes the first two phases: development and piloting. The development phase is described separately for the tool and training. Both elements of the intervention are joined in the piloting phase as a combined pilot study (see Figure 1 for an overview).

Figure 1. Schematic representation of the development process of the tool and training. HRQoL: health-related quality of life.


Design
Feedback sessions
Pilot testing
Web-based prediction tool
Communication skills training
Components: patient information entry, overall survival, toxicity, HRQoL and summary
Framework: R Shiny server with ggplot2 graphics
Components: e-learning, 2 face-to-face sessions
Focus group
6 (ex-)patients
Topic: clarity and usefulness of graphs
Individual interviews
5 individual patients
Topic: interpretation of graphs
Think-aloud sessions
3 medical oncologists
Topic: usability
Expert interviews
6 field experts
Topic: overall feedback
Combined application of training and tool
5 oncology fellows
Evaluation via semistructered interviews
Feedback tool and training
Real-life consultations
3 patients
Evaluation via semistructered interviews

The Web-Based Prediction Tool: Source

The web-based, physician-assisted prediction tool named *Source*, which contains visualizations of evidence-based, precise, and personalized outcome information, was developed using an iterative, user-centered approach. The tool was designed to be used by oncology health care providers for decision-making consultations. The Source tool, unlike other prediction tools such as *Predict* and *Adjuvant Online*, was not designed for unsupervised use by patients at home to prevent incorrect use, misunderstanding, and lack of emotional support.

First, prediction models were developed to ensure that the Source tool's treatment outcome information (survival, side effects and complications, and HRQoL) was evidence-based, precise, and personalized to the individual patient and tumor characteristics. Personalized predictions for survival using this tool are based on the SOURCE prediction models [45,46], which predict survival based on the individual patient and tumor characteristics and are regularly updated when new data become available. Depending on the tumor location, either nine variables (for gastric cancer) or 13 variables (for esophageal cancer) are required for predictions. These variables include age, tumor staging, and metastasis characteristics. The model output is the probability of survival up to 2 years following diagnosis and allows for the comparison of multiple treatments.

The side effects (toxicity) of chemotherapy treatment are based on TOXView meta-analyses [7]. These models establish the probability of adverse events such as nausea, alopecia, and neuropathy, stratified by mild or severe grade toxicity (according to the Common Terminology Criteria for Adverse Events [47]), for various chemotherapy regimens. These probabilities are not personalized to the individual characteristics and do not vary over time, as this was not possible with the available data.

Finally, predictions of HRQoL are available from a meta-analysis and describe the change in the EORTC QLQ-C30 (European Organisation for Research and Treatment of Cancer quality of life questionnaire C30) on the global health scale for best supportive care and chemotherapy in metastatic patients up to 6 months after diagnosis [10].

Next, these models and meta-analyses were used to visualize treatment outcomes. For the visualizations to be easy to understand for patients, a previous systematic literature review on visual risk communication was consulted [48]. Furthermore, the literature about usability and usability guidelines for web-based applications [49-52] and existing prediction tools, such as *Predict* and *Adjuvant Online*, were consulted [37,38]. On the basis of the literature, the first set of requirements for the tool was created according to the MoSCoW (Must Have, Should Have, Could Have, Won't Have) system [53]. This process resulted in the requirements listed in Textbox 1.

Textbox 1. Overview of the requirements of the web interface according to the MoSCoW (Must Have, Should Have, Could Have, Won't Have) system.

Must Have

1. After opening the tool, a data entry form is shown to enter the variables needed for the prediction models.
2. The data entry form is dynamic and shows only relevant variables.
3. Survival, adverse events, and health-related quality of life outcomes are displayed in their own tabs, and only one outcome is displayed at a time.
4. The outcomes are displayed graphically in a screen-filling image.

Should Have

1. The data entry form contains input validation to avoid mistakes during entry.
2. The data entry contains explanations of the variables.
3. The plots can be tailored to the patients' and physicians' preferences (eg, time frame and treatments to be compared).

Could Have

1. The tool's display language can be set to Dutch or English.
2. A textual summary accompanying the plots can be generated and printed so the patient can review the information at a later time.
3. A help function for physicians is incorporated.

A prototype of the web interface was created based on the literature guidelines and first requirements. The web-based tool was developed using the RShiny software (version 1.2.0; RStudio) supplemented with ggplot2 (version 3.2.1) to create graphs [54,55]. The creation, evaluation, and improvement of the tool followed an iterative user-centered design framework, where feedback was gathered from end users (patients and physicians) and experts. By iteratively updating the tool, we aim to provide improvements for the tool after each feedback session and avoid receiving the same feedback after each feedback round. A total of 4 feedback sessions were conducted from January 2018 to July 2018.

First, a face-to-face focus group was conducted with 6 patients with esophageal and gastric cancer and survivors from the Foundation for Patients with Cancer in the Digestive system after verbal informed consent was provided. The aim of this focus group was to obtain feedback on the tool in a group setting and promote discussions among the group members. One of the researchers acted as a moderator and presented the participants with each of the tool's graphs. Each displayed graph was accompanied by a short oral explanation, after which the participants were asked for their opinions. Feedback on the web interface and suggestions for improvement supported by multiple participants were used to create an improved version of the tool. The focus group session was audio-recorded and analyzed according to microinterlocutor analysis to systematically evaluate the participants' remarks [56].

In a second feedback round, 30-minute, semistructured, face-to-face interviews were conducted with individual patients with esophageal and gastric cancer. The main focus of these audio-recorded interviews was to determine whether the patient interpretation of the revised graphs was adequate. By conducting interviews with individual patients rather than in a group, the aim was to review patient interpretations without the influence of other patients. The interviews were conducted using a piloted script. A total of 5 outpatient participants were recruited by an oncologist at the Amsterdam University Medical Centers. Following the participants' informed consent, 2 researchers presented the tool to patients by using fictitious predictions of treatment outcomes. Patients were asked to interpret the presented graphs and describe their meaning to assess their understanding. Thereafter, the researcher provided the correct description of the graph, and the patients' subsequent feedback was gathered. Feedback was registered in a response matrix, including the frequency of different remarks, to establish which possible improvements could be implemented.

The third feedback round aimed to evaluate the usability of the tool when used by medical oncologists, and 3 medical oncologists at the Amsterdam University Medical Centers participated in individual face-to-face think-aloud sessions. After providing informed consent, they were asked to use the tool for two paper patient cases while stating out loud whatever came to mind. The cases described fictitious patients with esophageal or gastric cancer, including some of their clinical characteristics. Several tasks and questions about specific outcomes (eg, "What is the 1-year survival probability with best supportive care?" and "Which treatment has the best quality of life after 6 months?") were posed to guide the use of the tool by medical oncologists. At the beginning of the think-aloud session, a video explaining the think-aloud method [57] was presented and the participants were asked to complete a short practice exercise to ensure that they understood the think-aloud method before starting the task. After the think-aloud session, participants completed the System Usability Scale (SUS) to measure the ease of use and overall likeability of the web-based tool [52]. Both screen captures and audio recordings were registered during the think-aloud session. One of the researchers (FH) used both recordings to register whether the oncologists successfully completed the tasks, how many mouse clicks they used to complete a task, and which buttons they clicked on the web interface. The median SUS scores were calculated to provide a quantitative indication of usability.

In the fourth and last round, feedback from experts was gathered by conducting semistructured interviews with 6 researchers with expertise in patient-physician interaction, shared decision-making, risk communication, medical informatics, and clinical decision support software. The experts were presented with a walkthrough of the tool and its options. Interviews were recorded and summarized to determine which possible improvements were brought forward.

The Source Supportive Communication Skills Training

Overview

Communication skills training was developed to educate physicians on informing patients with cancer in a treatment decision-making consultation using the Source tool. Due to the complexity of the skills needed, it was important to specify clear learning goals. As stated in complex learning theory, when training complex skills, the desired learning outcomes must address the following domains: knowledge, attitudes, and skills [58-60]. The training aimed for physicians to be able to name the most important tips and tricks for adequately informing patients on treatment outcomes and communicating treatment risks and benefits (knowledge). Furthermore, the training aimed for physicians to have a positive outlook on using numbers to inform patients about treatment outcomes and their ability to inform patients in an evidence-based, precise, personalized, and tailored manner (attitude). Moreover, the training aimed for physicians to be able to use the Source tool and to incorporate the tool to inform patients during consultations (skills). Finally, the training aimed to increase physicians' ability to provide information tailored to patients' informational needs and level of understanding (skills). A team (n=5) of experts in medical communication and psycho-oncology and experienced trainers in medical communication discussed the context and content of the training and set learning objectives. In addition, the literature on training and shared decision-making frameworks was reviewed.

As physicians value time-efficient and flexible training [16], the training was designed as blended learning, encompassing preparatory e-learning and a face-to-face component. The 4-step shared decision-making model proposed by Stiggelbout et al [12] was used as a framework. This model distinguishes the following four essential steps for shared decision-making: (1) setting the agenda, (2) informing about treatment options, (3) exploring patients' values, and (4) making a decision in agreement [61]. The outline of the training was based on previous communication skills training for skills in shared decision-making, as designed for and proven to be effective in the CHOICE (Choosing Treatment Together in Cancer at the End of Life) trial [61] and the literature on the guidelines for effective communication skills training [42,62]. The focus of the Source training is the second step of this model, that is, informing patients about treatment options and the pros and cons thereof.

e-Learning

First, e-learning was targeted at summarizing the evidence base for effective information provision and providing physicians with tips and tricks for clinical practice. to this end, we consulted the literature related to theories, evidence, and guidelines on the provision of information in medical practice [12,63-67]. The assembled literature and theories were summarized into short chapters, each covering a different subtopic. The expert team discussed the scripts in these chapters to obtain a consensus on the frameworks and models used. Interactive elements, such as exercises, were added to the e-learning to enable the learner to actively process the information. Second, the e-learning aimed to introduce the Source tool, thereby addressing the use and functionalities of the tool and the underlying prediction models.

An earlier study concluded that physicians value both visual attractiveness and variation between learning activities in e-learning [16]; therefore, the layout, animations, and videos were developed in cooperation with a small visual design company, Public Cinema.

Face-to-Face Sessions

Face-to-face sessions were developed based on previous experience in developing and evaluating communication skills training in oncology [42,61,68]. The most important recommendations from these earlier studies were to role-play with an actor to practice the lessons learned during the training and provide the trainee with personal feedback [42]. Development took place in multiple sessions with the expert team. The basic assumptions for effective information provision, as incorporated in the e-learning, served as a starting point for the training content. Derived ideas were written down and discussed to create a training script and a supportive PowerPoint presentation. The casuistry for the training actors was developed together with a clinical expert (HWMVL). These multiple development sessions led to a conceptual version of the training.

Pilot Study Tool and Training

A pilot study was conducted from December 2018 to March 2019 to test both the tool and training in a real-life setting. As this pilot study targeted patients with advanced disease only, we included medical oncologists and metastatic cancer patients as study participants. The pilot study was evaluated by the Medical Ethics Review Committee of the Academic Medical Center Amsterdam (reference number: W18_278). In total, 5 medical oncology fellows (2 men and 3 women) from two university medical centers were invited to use the tool and trained according to the concept training format. After completion of the training, participating fellows were individually interviewed via telephone in a semistructured manner to gather feedback to improve both the Source tool and training. For the tool, the focus of the feedback was on opinions and experiences regarding usability and willingness to use the tool. Regarding training, feedbacks on the different components, training as a whole, and perceived utility were collected.

In addition, an experienced medical oncologist (HWMVL) conducted three treatment decision-making consultations with outpatients using the Source tool for information provision and in line with the training principles. These consultations were recorded on video after obtaining written informed consent from the patients and oncologists. To comply with ethical standards and according to the training, only information that the patient wanted to receive was disclosed to the patient. One-on-one semistructured interviews were conducted with the 3 patients by one of the researchers (LFVDW) to gather their experiences with the physician's outcome information and the use of the Source tool.

Results

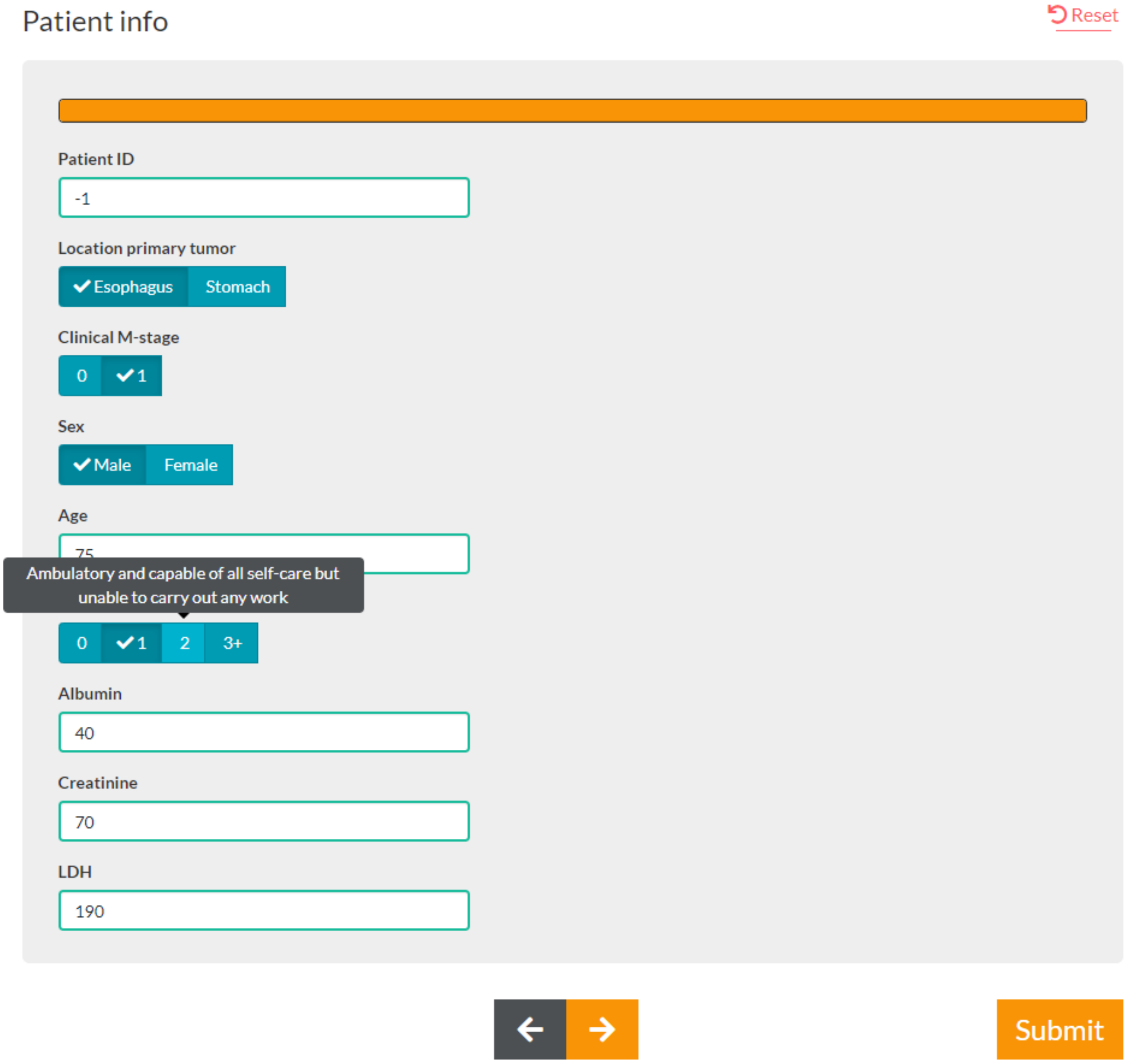
The Web-Based Prediction Tool: Source

A prototype web interface was created based on the findings of a systematic review of the effects of different types of risk communication on patients with cancer [48]. Following this review, we decided to use clear and precise risk information (eg, percentages or frequencies) and simple graphs with a limited amount of information displayed. As the review did not yield consistent guidelines on which types of graphs to use, it was decided to visualize the outcomes in multiple ways. In this way, graphs can be used according to the preferences of individual patients, and the amount and presentation format of the information displayed can be tailored to their needs and preferences. The resulting RShiny web interface runs on an x64 Linux server (version 3.10.0).

Source Tool Components

The final tool, Source 1.103, contains five main components. The first component, the patient information entry component, allows the oncologist to enter the patient characteristics necessary for the prediction models and meta-analyses, using supporting information, such as the definitions of TNM variables (Figure 2).

Figure 2. A screenshot of patient data entry. The data entry screen displays the fields that are necessary for the prediction models and meta-analyses. Additional information on variables such as the World Health Organization performance status is provided with a mouse-over.



Patient info

Reset

Patient ID
-1

Location primary tumor
✓Esophagus Stomach

Clinical M-stage
0 ✓1

Sex
✓Male Female

Age
75

Ambulatory and capable of all self-care but unable to carry out any work

0 ✓1 2 3+

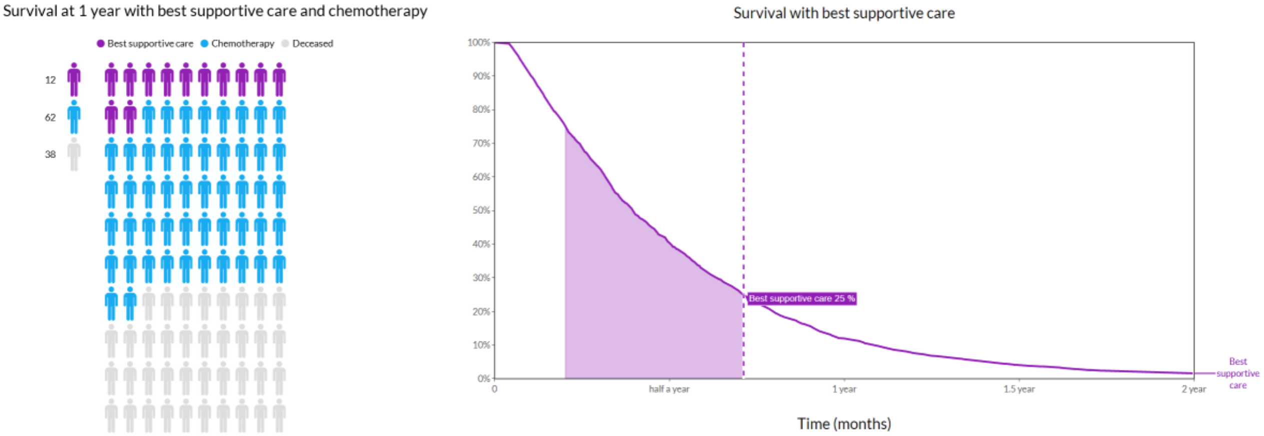
Albumin
40

Creatinine
70

LDH
190

The survival component outcome was visualized in two ways: an icon array displaying the survival probability at a given point in time by coloring a subset of 100 figures and a Kaplan-Meier curve (line graph) displaying the survival probability over time (Figure 3). The survival component incorporates the possibility of switching between the two presentation formats. From the options menu, it is possible to select specific treatments for comparison (eg, best supportive care and chemotherapy) and change the time frame of the prediction (from 6 to 24 months) to show a per treatment CI and three survival scenarios (indicating the best-case scenario comprising the top 25%, the worst-case scenario comprising the bottom 25%, and the typical outcome comprising the middle 50% [63,64]).

Figure 3. A screenshot of survival graphs. On the left, a pictograph displaying the predicted survival for best supportive care and chemotherapy after 1 year is shown. On the right, the Kaplan-Meier curve for the best supportive care is shown. The optional shaded area displays the so-called typical outcome scenario (with survival ranging from 25% to 75%).



The side effects component displays bar charts for various toxicities (Figure 4), as the meta-analysis provided static probabilities for each of the adverse events [7]. Each side effect was visualized by two stacked bars, one for mild side effects and one for severe side effects on the Common Terminology Criteria for Adverse Events scale [69]. The side effects of multiple chemotherapy regimens can also be compared. To avoid information overload, only the three most frequently occurring side effects are shown initially, although it is possible to select all side effects.

HRQoL is displayed in a line graph and shows the EORTC QLQ-C30 global health score over time (Figure 5). There are options to compare HRQoL in best supportive care with chemotherapy and display a CI and reference value (obtained from the European Organisation for Research and Treatment of Cancer reference values manual [70]). The final component is a summary that can be printed as a handout for the patient or saved as a PDF file. This feature enables physicians to show the aforementioned graphs accompanied by an explanatory text. This text is dynamically generated using the selected treatment data and explains the content of the graphs.

Figure 4. A screenshot of the side effects bar chart. This displays the three most commonly occurring toxicities for both 5FU and CAPOX. The darker bars indicate severe toxicities, and the lighter bars indicate mild toxicities. 5FU: 5-Fluorouracil; CAPOX: capecitabine combined with oxaliplatin.

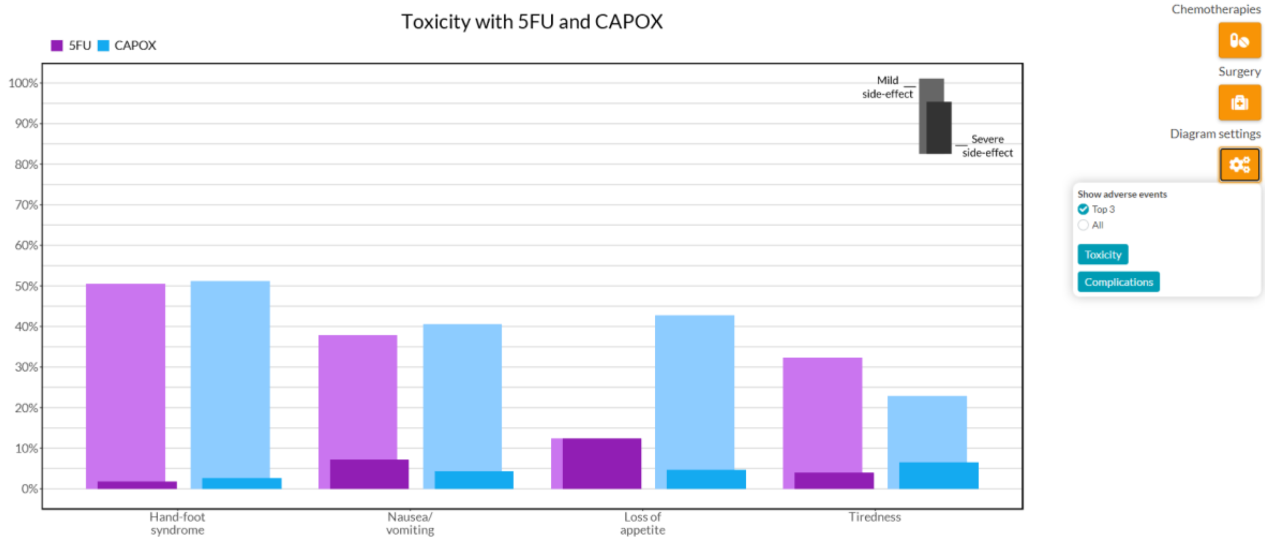
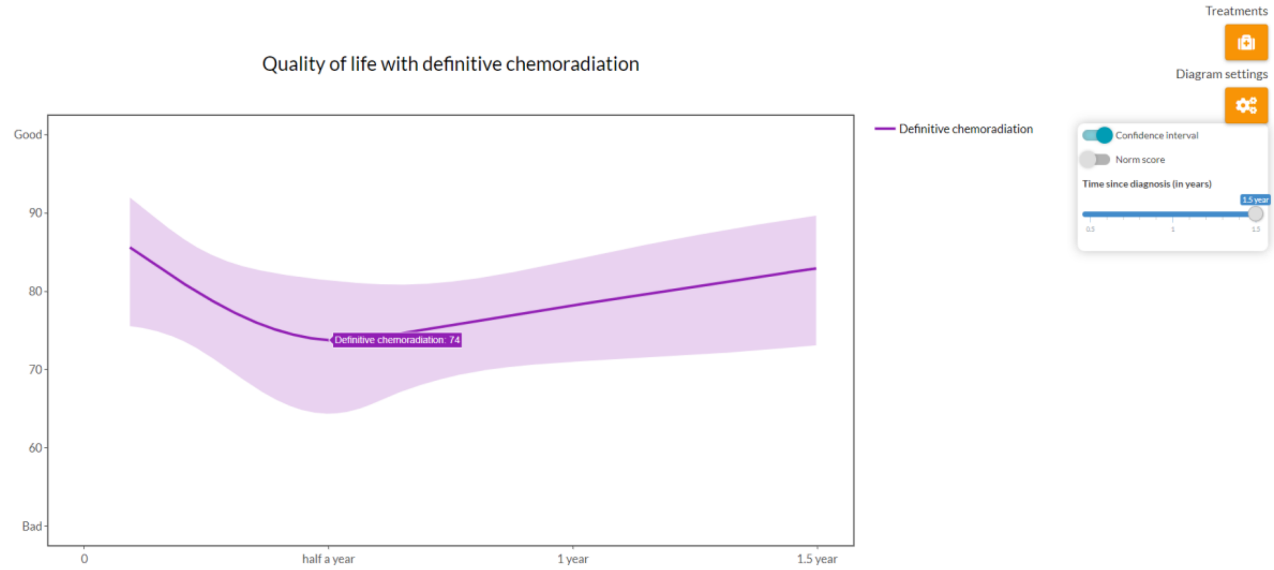


Figure 5. A screenshot of the HRQoL (health-related quality of life) line graph. The graph displays the HRQoL following definitive chemoradiation. The shaded area optionally displays the CI of the HRQoL line.


Quality of life with definitive chemoradiation
Good
90
80
70
60
Bad
0
half a year
1 year
1.5 year
Definitive chemoradiation: 74
Definitive chemoradiation
Treatments
Diagram settings
0.5
1
1.5

● Confidence interval
○ Norm score
Time since diagnosis (in years): 1.5 year

Evaluation Round Feedback

The four evaluation rounds of the tool resulted in several minor visual and functional adjustments to the web interface, as described in Multimedia Appendix 1. Major adjustments to the tool resulting from the gathered feedback were mostly adjustments regarding usability, such as increasing the font size and positioning of the legend. For the survival outcome, the icon array was found to be the most comprehensible, whereas the line graph provided the most insight for survival over time. Therefore, it was decided to keep both formats for the survival outcome, as the graphs supplemented each other. The line graph also remained in the tool as it incorporated the scenario's functionality (indicating worst-case, typical, and best-case survival), a feature that was found important by most patients. For side effects, it was found that the patients did not correctly interpret the meaning of the stacked bar charts. The bars were changed into two nonstacked bars with 90% overlap to display mild and severe adverse events for the same side effects to increase clarity on the meaning of the graph (Figure 4). Furthermore, it was decided to remove bar charts as a display option for HRQoL data, as both patients and oncologists found the graph unclear and wanted the data to be displayed over time, as HRQoL may increase and decrease over time. Showing the predictions at a single time point may therefore not provide sufficient insight. In the final design, various options are available to personalize the displayed graphs. Regarding the usability of the tool, it received a median SUS score of 90.0 out of 3 ratings (above the *Excellent* threshold of 80.3 points [52]) during the think-aloud sessions.

In the pilot study, 4 of the 5 oncology fellows participated in an evaluative semistructured interview following the training. The fifth fellow did not respond to repeated invitations. All 4 oncology fellows reported that the tool was highly usable overall. Some minor suggestions were provided for improving the display of certain options, graphs, and buttons. Of the 4 oncologists, 3 reported that they would use the tool in their clinical practice. They especially valued the personalized nature of the tool's predictions and the clear and easy-to-understand visualizations for patients. Furthermore, the inclusion of HRQoL data and the option to print a summary for the patient were found useful. The fourth fellow would like the prediction models to be further developed before using the tool. For instance, during the pilot training and interviews, several critical remarks on the prediction models were expressed, such as the lack of World Health Organization performance status as a predictor of overall survival. These comments likely reflect a possible lack of trust in the underlying models and analyses of the tool. Another barrier to using the tool that one of the fellows addressed was the fear of emotionally confronting the patient with the exact numbers. This fellow did indicate that this fear was overcome through the tailoring skills that he had acquired during the training.

The Source Supportive Communication Skills Training

The e-Learning Module

A short and to-the-point e-learning module was developed to provide an overview of the theoretical background on the provision of information in a treatment decision context and introduce a web-based prediction tool. The e-learning module starts with a short peer endorsement video of the training of physicians that discusses the Source tool. Subsequently, physicians can navigate through 4 chapters. The first chapter provides an overview of the principles of effective information provision in the context of a shared decision. The second chapter introduces the physicians to the Source tool by presenting them with a tailor-made instructional video of its use and functions. Furthermore, a summary of the tool's models and their underlying data is provided. The third chapter provides an overview of tips and tricks for informing patients about the risks and benefits of treatment (Multimedia Appendix 2). The final chapter consists of a short and practical summary of key take-home messages. In all chapters, textual information and short assignments are alternated by instructional videos and animated knowledge clips. A simple and appealing visual design is applied.

Face-to-Face Training Sessions

The face-to-face component of the training consisted of two group sessions of 3.5 hours each, provided by an experienced trainer, with approximately 2 to 3 weeks in between to facilitate intermediate practice. The sessions were aimed at small groups of 2 to 6 participants, as this approach enables every participant to practice and receive personal feedback. Such a setting can also promote interactivity [71]. Both sessions involved individual role-play exercises with a professional actor in which feedback was provided by the trainer, the actor, and peers to learn additional skills [62]. Furthermore, group discussions were stimulated and led by the trainers. This approach was used to encourage physicians to discuss their attitudes toward using numbers and the tool in the context of the provision of information to patients [62,72].

The first session covered the skills of setting the agenda of the decision-making consultation, introducing the tool and informing patients on survival outcomes. Physicians were asked to practice separate parts of the consultation while receiving feedback from other physicians and the trainer. Physicians were instructed to use the acquired knowledge and skills during their outpatient consultations before the next training session. The second session allowed for the repetition of issues addressed during the first session and sharing experiences of applying the lessons learned of the first session in clinical practice. Next, skills were addressed, again with role-play and feedback, to inform patients about treatment outcomes in terms of side effects and complications and HRQoL. This session also addressed how to conclude decision-making consultations.

In both sessions, tailoring of the amount and type of outcome information to specific patients played a significant role in both role-play practice and group discussions. Tips and tricks were discussed regarding how to determine an individual's information needs and wants, how to fit these needs with the informational needs of the physician, and whether and how the tool could contribute to tailored information giving.

The total duration of the blended learning was 7.5 study hours, which consisted of 0.5 hours of e-learning and 7 hours of face-to-face training.

Feedback and Major Adjustments

Medical Oncology Fellows

Fellows reported that they enjoyed participating in the training and specifically valued personal coaching and practical tips. Furthermore, fellows appreciated the trainers and actors. In their opinion, there was a good balance between the information provided by trainers and practical exercises. Fellows especially appreciated the feedback during the training from the actor and trainer on their role-play with the actor. Overall, the training was described as useful, and specific improvements were suggested.

A point of improvement that was brought up was the substantial time investment in the training. In particular, the pace of the e-learning and instructional video was considered too slow. Furthermore, the timing of the face-to-face sessions following a day of work was considered inconvenient. The most important adjustments to the training as a result of the fellows' feedback were related to accelerating the pace and adding an individual booster session to the training in which the physician could receive personal feedback from one of the trainers on a full, recorded consultation.

Patients With Metastatic Cancer

Despite their emotions on the subject, 2 out of the 3 patients were willing to participate in a short interview about their experiences with the consultation. Both appreciated the use of the Source tool and the physicians' explanations of possible treatment outcomes. Patients expressed differences in their experiences regarding the amount of information about treatment options and outcomes. Although one patient reported being satisfied with the amount of information, the other indicated that the amount of information was too extensive for him to memorize it all. He needed a printed summary of the tool for support. Both patients mentioned their struggle with the meaning of risk or benefit for themselves, as great uncertainty remains about their own future, despite the information provided.

Discussion

Findings

Our study shows the iterative development and pilot testing of the Source tool and training. This combined intervention was developed using scientific evidence and input from physicians, patients, and experts. This process resulted in the first web-based prediction tool to inform patients with esophagogastric cancer during consultations on survival, side effects, and HRQoL of different treatment options. Furthermore, we created a supporting training to teach physicians the communication skills needed to use the tool and to provide patients with information in an evidence-based, precise, personalized, and tailored manner. Preliminary evaluation results are promising and provide valuable input for further development and testing of both elements.

Both the tool and training were valued by participating physicians and patients. Physicians especially appreciated the practical approach of the training; the multiple practice opportunities and personal feedback helped them use the tool. Nevertheless, despite their positive attitudes toward the tool and training, old habits could stand in the way of using the tool and may impede the use of learned communication techniques in clinical practice. Behavioral change theories show that many factors can contribute to but also stand in the way of learning new behaviors. Resistance could, for instance, arise as a result of a different expected outcome of the tool or because of a low tolerance for change [43]. The *transfer* of training describes the possible behavioral change resulting of an educational intervention such as training. From the literature, we know that although certain trainee characteristics (such as the perceived utility of the training) and training design factors (such as a realistic training environment) can promote the transfer of training, they are also strongly influenced by characteristics of the work environment, such as situational cues (eg, social support from peers or supervisors) or consequences (eg, negative or strong emotional reactions from patients) [73]. These characteristics can be difficult to control in the setting of everyday hospital care. However, concerning the training design factors, the distribution of the training sessions might be an important factor contributing to the transfer of this tool and training for daily clinical practice. Indeed, the so-called *spacing* between the two face-to-face sessions and the booster session of the training might help increase task performance [74-79].

During the pilot training, it was noted that in some cases, the fellows lacked trust in the prediction models used in the tool. Further steps were taken to increase the physicians' trust. For instance, details about the underlying data and publications were added to the tool to provide more information about the methodology and sources. Model updates (such as the 2020 version of the survival model), which increases the model's performance and sample size and includes, for instance, World Health Organization performance status as a predictor, may also increase trust in the tool [39]. Finally, external validation of the models can also generate trust in the validity and applicability of the tool [80].

The use of the Source tool could also be influenced by the application's usability and how well the tool solves the patient-informing problem as perceived by patients and physicians. Therefore, iterative usability testing is necessary to achieve an acceptable level of usability. As the number of testing rounds in this study is limited compared with other studies [37,38], the tool's usability may be further improved. This issue will be an ongoing point of attention that we will address during future testing and development of the tool. From a patient's perspective, some uncertainties regarding treatment outcomes may be reduced during the consultation, whereas other uncertainties remain. For example, the tool might support patients in participating in shared decision-making, but active participation in this difficult choice might also overwhelm them [81]. These issues can be addressed in medical education and training by dealing with a broad spectrum of patient uncertainty.

On the basis of our experiences, we can provide several recommendations to aid future research in creating and evaluating web-based prediction tools with training. First, we advise involving end users, such as patients and physicians, in the early stages of development. Assumptions and implementations are often made from the perspective of developers, which may not coincide with the needs and wishes of patients or physicians. By evaluating at an early stage, it is possible to adjust the tool and training, and subsequent improvements can be implemented more seamlessly. Although not formally evaluated in this study, user research to investigate patients' and physicians' ideas and expectations regarding such a tool could also contribute to the usability and adoption of the tool in clinical practice. Second, evaluation by physicians may be complicated because of their busy schedules. We recommend making the feedback rounds with physicians as short as possible, planning them sufficiently in advance, and having them take place on education days. Third, as insights on data visualization and risk communication may change constantly, we recommend facilitating ongoing updates of a designed prediction tool. We also suggest that future research should use the current state of the art when designing a new tool or training. Fourth, we advise that communication on outcomes with subjective interpretations, such as HRQoL, deserves a more prominent place in communication skills training. We noticed in our training that physicians often had trouble explaining outcomes such as HRQoL. Finally, it was observed that end users sometimes had conflicting opinions regarding improvements to the tool or training. As it is not possible to cater to everyone's wishes, we recommend weighing the pros and cons of suggestions and deciding whether a personalization option will be implemented (such as displaying survival as a pictograph and a line graph) or whether a single option will be implemented (such as the background color of the web interface).

Most of the patients and physicians who participated in this study agreed that the tool and training added value to clinical practice. However, bias may have played a role in the evaluation of the tool and training, as the evaluation only partly took place in clinical practice. To investigate this potential bias and the extent to which the combined tool and training aid in information provision in the context of treatment decision-making, we are currently performing the third phase of the Medical Research Council framework: evaluation, a multicenter effect study (registered under NCT04232735, the SOURCE trial). In this stepped-wedge trial, physicians receive training and use the Source tool both in simulated patient assessments and with outpatients. The effect that the intervention has on the outcome information provided by oncology physicians was quantitatively investigated by recording these consultations before and after the intervention and analyzing physicians' outcome-related remarks. The primary outcome of this study is the provision of precise outcome information; secondary outcomes include the amount of tailoring to the information needs of patients, the patients' own knowledge and opinions on the communicated outcome information, and the influence that the consultation has on patients' emotions. As trust in the SOURCE models was found to be a potential barrier to using the tool in the pilot study, physicians' trust in the models will be closely monitored and specifically addressed during the trial. The models included in the Source tool are being continuously improved and updated, in part, to address these issues.

In this trial, both palliative and curative patients will participate, and models aimed at potentially curable patients (cM0) will be added to the Source tool. The survival models are based on the 2020 version of the models, which include updated palliative prediction models and newly developed curative prediction models for both esophageal and gastric cancer [39]. The HRQoL model for curatively treated patients originates from a systematic review and meta-analysis, and treatment side effects for curatively treated patients were provided by the COMplot study [82,83]. The addition of these models to the Source tool enables evidence-based and precise information personalized to the individual's characteristics in the full spectrum of patients with esophagogastric cancer. As the tool is currently being tested in a trial, access is currently restricted to trial participants. However, after the conclusion of the trial, the tool will be freely available in both Dutch and English, enabling the use of the Source tool in clinical practice [84].

Conclusions

We developed and evaluated a web-based tool and training to inform patients with esophageal or gastric cancer regarding treatment outcomes. Through evaluation and a pilot study, patients and physicians indicated the added value of the tool and training, and both were improved based on their feedback. The tool and training are currently being evaluated in a multicenter trial to determine their added value in clinical practice.

Acknowledgments

Financial support for this study was provided entirely by a grant from the Dutch Cancer Society (grant UVA 2014-7000). The funding agreement ensured the authors' independence in designing the study, interpreting the data, and writing and publishing the report. The funder had no role in the design and conduct of the study; the collection, management, analysis, and interpretation of the data; the preparation, review, or approval of the manuscript; or the decision to submit the manuscript for publication.

Conflicts of Interest

The authors developed and evaluated the web-based tool Source and the accompanying training described in this study. HWMVL reports grants from the Dutch Cancer Society during the conduct of the study and personal fees from Bristol Myers Squibb, Lilly, and Nordic Pharma; and grants and nonfinancial support from Bayer, Bristol Myers Squibb, Celgene, Jansen, Lilly, Nordic Pharma, Philips, Roche, and Servier outside this work. In addition, HWMVL has a consultant or advisory role with Bristol Myers Squibb, Lilly, MSD, Nordic Pharma, and Servier.

Multimedia Appendix 1
Detailed adaptations to the Source tool during the modeling phase.
[DOCX File , 19 KB - jmir_v23i8e27824_app1.docx]

Multimedia Appendix 2
Recommendations on risk communication displayed in the e-learning training.
[DOCX File , 15 KB - jmir_v23i8e27824_app2.docx]

References

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 2015 Mar 1;136(5):359-386. [doi: 10.1002/ijc.29210] [Medline: 25220842]
2. Badgwell B, Das P, Ajani J. Treatment of localized gastric and gastroesophageal adenocarcinoma: the role of accurate staging and preoperative therapy. J Hematol Oncol 2017 Aug 15;10(1):149 [FREE Full text] [doi: 10.1186/s13045-017-0517-9] [Medline: 28810883]
3. van Hagen P, Hulshof MC, van Lanschot JJ, Steyerberg EW, van Berge Henegouwen MI, Wijnhoven BP, CROSS Group. Preoperative chemoradiotherapy for esophageal or junctional cancer. N Engl J Med 2012 May 31;366(22):2074-2084. [doi: 10.1056/NEJMoa1112088] [Medline: 22646630]
4. Smyth EC, Verheij M, Allum W, Cunningham D, Cervantes A, Arnold D, ESMO Guidelines Committee. Gastric cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol 2016 Sep;27(suppl 5):38-49 [FREE Full text] [doi: 10.1093/annonc/mdw350] [Medline: 27664260]
5. Lordick F, Mariette C, Haustermans K, Obermannová R, Arnold D, ESMO Guidelines Committee. Oesophageal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol 2016 Sep;27(suppl 5):50-57. [doi: 10.1093/annonc/mdw329] [Medline: 27664261]
6. van den Ende T, Ter Veer E, Machiels M, Mali RM, Nijenhuis FA, de Waal L, et al. The efficacy and safety of (Neo) adjuvant therapy for gastric cancer: a network meta-analysis. Cancers (Basel) 2019 Jan 11;11(1):80 [FREE Full text] [doi: 10.3390/cancers11010080] [Medline: 30641964]
7. Ngai LL, ter Veer E, van den Boorn HG, van Herk EH, van Kleef JJ, van Oijen MG, et al. TOXview: a novel graphical presentation of cancer treatment toxicity profiles. Acta Oncol 2019 Aug;58(8):1138-1148. [doi: 10.1080/0284186X.2019.1601256] [Medline: 31017020]
8. ter Veer E, Mohammad NH, van Valkenhoef G, Ngai LL, Mali RM, Anderegg MC, et al. The efficacy and safety of first-line chemotherapy in advanced esophagogastric cancer: a network meta-analysis. J Natl Cancer Inst 2016 Oct;108(10):djw166. [doi: 10.1093/jnci/djw166] [Medline: 27576566]
9. ter Veer E, Mohammad NH, van Valkenhoef G, Ngai LL, Mali RM, van Oijen MG, et al. Second- and third-line systemic therapy in patients with advanced esophagogastric cancer: a systematic review of the literature. Cancer Metastasis Rev 2016 Sep;35(3):439-456 [FREE Full text] [doi: 10.1007/s10555-016-9632-2] [Medline: 27417221]

10. van Kleef JJ, ter Veer E, van den Boorn HG, Schokker S, Ngai LL, Prins MJ, et al. Quality of life during palliative systemic therapy for esophagogastric cancer: systematic review and meta-analysis. J Natl Cancer Inst 2020 Jan 01;112(1):12-29. [doi: 10.1093/jnci/djz133] [Medline: 31251346]
11. O'Connor AM, Légaré F, Stacey D. Risk communication in practice: the contribution of decision aids. Br Med J 2003 Sep 27;327(7417):736-740 [FREE Full text] [doi: 10.1136/bmj.327.7417.736] [Medline: 14512487]
12. Stiggelbout AM, Pieterse AH, De Haes JC. Shared decision making: Concepts, evidence, and practice. Patient Educ Couns 2015 Oct;98(10):1172-1179. [doi: 10.1016/j.pec.2015.06.022] [Medline: 26215573]
13. van den Ende T, ter Veer E, Mali RM, van Berge Henegouwen MI, Hulshof MC, van Oijen MG, et al. Prognostic and predictive factors for the curative treatment of esophageal and gastric cancer in randomized controlled trials: a systematic review and meta-analysis. Cancers (Basel) 2019 Apr 12;11(4):530 [FREE Full text] [doi: 10.3390/cancers11040530] [Medline: 31013858]
14. ter Veer E, van Kleef JJ, Schokker S, van der Woude SO, Laarman M, Mohammad NH, et al. Prognostic and predictive factors for overall survival in metastatic oesophagogastric cancer: a systematic review and meta-analysis. Eur J Cancer 2018 Nov;103:214-226. [doi: 10.1016/j.ejca.2018.07.132] [Medline: 30268922]
15. Dugdale DC, Epstein R, Pantilat SZ. Time and the patient-physician relationship. J Gen Intern Med 1999 Jan;14 Suppl 1:34-40 [FREE Full text] [doi: 10.1046/j.1525-1497.1999.00263.x] [Medline: 9933493]
16. Stuij SM, Labrie NH, van Dulmen S, Kersten MJ, Christoph N, Hulsman RL, INSTRUCT project group. Developing a digital communication training tool on information-provision in oncology: uncovering learning needs and training preferences. BMC Med Educ 2018 Sep 24;18(1):220 [FREE Full text] [doi: 10.1186/s12909-018-1308-x] [Medline: 30249221]
17. Politi MC, Han PK, Col NF. Communicating the uncertainty of harms and benefits of medical interventions. Med Decis Making 2007;27(5):681-695. [doi: 10.1177/0272989X07307270] [Medline: 17873256]
18. Visser LN, Tollenaar MS, van Doornen LJ, de Haes HC, Smets EM. Does silence speak louder than words? The impact of oncologists' emotion-oriented communication on analogue patients' information recall and emotional stress. Patient Educ Couns 2019 Jan;102(1):43-52. [doi: 10.1016/j.pec.2018.08.032] [Medline: 30224275]
19. Kessels RP. Patients' memory for medical information. J R Soc Med 2003 May;96(5):219-222 [FREE Full text] [doi: 10.1258/jrsm.96.5.219] [Medline: 12724430]
20. Clayton JM, Adler JL, O'Callaghan A, Martin P, Hynson J, Butow PN, et al. Intensive communication skills teaching for specialist training in palliative medicine: development and evaluation of an experiential workshop. J Palliat Med 2012 May;15(5):585-591. [doi: 10.1089/jpm.2011.0292] [Medline: 22433021]
21. Politi MC, Légaré F. Physicians' reactions to uncertainty in the context of shared decision making. Patient Educ Couns 2010 Aug;80(2):155-157. [doi: 10.1016/j.pec.2009.10.030] [Medline: 19948387]
22. McIntosh J. Processes of communication, information seeking and control associated with cancer: a selective review of the literature. Soc Sci Med 1974 Apr;8(4):167-187. [doi: 10.1016/0037-7856(74)90107-3] [Medline: 4600056]
23. Kodish E, Post SG. Oncology and hope. J Clin Oncol 1995 Jul;13(7):1817. [doi: 10.1200/JCO.1995.13.7.1817] [Medline: 7602370]
24. Good M. The practice of biomedicine and the discourse on hope. In: Pfleiderer B, Bibeau G, editors. Anthropologies of Medicine. Switzerland: Springer; 1991:121-135.
25. Daugherty CK, Hlubocky FJ. What are terminally ill cancer patients told about their expected deaths? A study of cancer physicians' self-reports of prognosis disclosure. J Clin Oncol 2008 Dec 20;26(36):5988-5993 [FREE Full text] [doi: 10.1200/JCO.2008.17.2221] [Medline: 19029419]
26. Douma KF, Koning CC, de Haes HC, Zandbelt LC, Stalpers LJ, Smets EM. Do radiation oncologists tailor information to patients needs? And, if so, does it affect patients? Acta Oncol 2012 Apr;51(4):512-520. [doi: 10.3109/0284186X.2012.665476] [Medline: 22414096]
27. Oerlemans S, Husson O, Mols F, Poortmans P, Roerdink H, Daniels LA, et al. Perceived information provision and satisfaction among lymphoma and multiple myeloma survivors - results from a Dutch population-based study. Ann Hematol 2012 Oct;91(10):1587-1595 [FREE Full text] [doi: 10.1007/s00277-012-1495-1] [Medline: 22638756]
28. Hack TF, Degner LF, Parker PA. The communication goals and needs of cancer patients: a review. Psychooncology 2005 Oct;14(10):831-846. [doi: 10.1002/pon.949] [Medline: 16200519]
29. Puts MT, Papoutsis A, Springall E, Tourangeau AE. A systematic review of unmet needs of newly diagnosed older cancer patients undergoing active cancer treatment. Support Care Cancer 2012 Jul;20(7):1377-1394. [doi: 10.1007/s00520-012-1450-7] [Medline: 22476399]
30. Parameswaran R, McNair A, Avery KN, Berrisford RG, Wajed SA, Sprangers MA, et al. The role of health-related quality of life outcomes in clinical decision making in surgery for esophageal cancer: a systematic review. Ann Surg Oncol 2008 Sep;15(9):2372-2379. [doi: 10.1245/s10434-008-0042-8] [Medline: 18626719]
31. Koedoot CG, Oort FJ, de Haan RJ, Bakker PJ, de Graeff A, de Haes JC. The content and amount of information given by medical oncologists when telling patients with advanced cancer what their treatment options are. palliative chemotherapy and watchful-waiting. Eur J Cancer 2004 Jan;40(2):225-235. [doi: 10.1016/j.ejca.2003.10.008] [Medline: 14728937]

32. Henselmans I, Jacobs M, van Berge Henegouwen MI, de Haes HC, Sprangers MA, Smets EM. Postoperative information needs and communication barriers of esophageal cancer patients. Patient Educ Couns 2012 Jul;88(1):138-146 [FREE Full text] [doi: 10.1016/j.pec.2011.12.004] [Medline: 22244819]
33. Henselmans I, Smets EM, Han PK, de Haes HC, Laarhoven HW. How long do I have? Observational study on communication about life expectancy with advanced cancer patients. Patient Educ Couns 2017 Oct;100(10):1820-1827. [doi: 10.1016/j.pec.2017.05.012] [Medline: 28511804]
34. Zandbelt LC, Smets EM, Oort FJ, Godfried MH, de Haes HC. Satisfaction with the outpatient encounter: a comparison of patients' and physicians' views. J Gen Intern Med 2004 Nov;19(11):1088-1095 [FREE Full text] [doi: 10.1111/j.1525-1497.2004.30420.x] [Medline: 15566437]
35. Jacobs M. Supporting patients in obtaining and oncologists in providing evidence-based health-related quality of life information prior to and after esophageal cancer surgery. In: PhD thesis. Amsterdam: University of Amsterdam; 2015:1-235.
36. Hodgkinson K, Butow P, Hunt GE, Pendlebury S, Hobbs KM, Lo SK, et al. The development and evaluation of a measure to assess cancer survivors' unmet supportive care needs: the CaSUN (Cancer Survivors' Unmet Needs measure). Psychooncology 2007 Sep;16(9):796-804. [doi: 10.1002/pon.1137] [Medline: 17177268]
37. Ravdin PM, Siminoff LA, Davis GJ, Mercer MB, Hewlett J, Gerson N, et al. Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. J Clin Oncol 2001 Feb 15;19(4):980-991. [doi: 10.1200/JCO.2001.19.4.980] [Medline: 11181660]
38. Wishart GC, Azzato EM, Greenberg DC, Rashbass J, Kearins O, Lawrence G, et al. PREDICT: a new UK prognostic model that predicts survival following surgery for invasive breast cancer. Breast Cancer Res 2010;12(1):R1. [doi: 10.1186/bcr2464] [Medline: 20053270]
39. van den Boorn HG, Engelhardt EG, van Kleef J, Sprangers MA, van Oijen MG, Abu-Hanna A, et al. Prediction models for patients with esophageal or gastric cancer: a systematic review and meta-analysis. PLoS One 2018;13(2):e0192310 [FREE Full text] [doi: 10.1371/journal.pone.0192310] [Medline: 29420636]
40. Farmer G, Pearson G, Skylark W, Freeman A, Spiegelhalter D. Making prognostic algorithms useful in shared decision-making: patients and clinicians requirements for the Predict: Breast Cancer interface. medRxiv 2020:A. [doi: 10.1101/2020.11.16.20232348]
41. Butler L, Degner L, Baile W, Landry M, SCRN Communication Team. Developing communication competency in the context of cancer: a critical interpretive analysis of provider training programs. Psychooncology 2005 Oct;14(10):861-873. [doi: 10.1002/pon.948] [Medline: 16200525]
42. Bos-van den Hoek DW, Visser LN, Brown RF, Smets EM, Henselmans I. Communication skills training for healthcare professionals in oncology over the past decade: a systematic review of reviews. Curr Opin Support Palliat Care 2019 Mar;13(1):33-45. [doi: 10.1097/SPC.0000000000000409] [Medline: 30562180]
43. Kotter JP, Schlesinger LA. Choosing strategies for change. Harv Bus Rev 1979;57(2):106-114. [Medline: 10240501]
44. 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]
45. van den Boorn HG, Abu-Hanna A, Mohammad NH, Hulshof MC, Gisbertz SS, Klarenbeek BR, et al. SOURCE: prediction models for overall survival in patients with metastatic and potentially curable esophageal and gastric cancer. J Natl Compr Canc Netw 2021 Apr;19(4):403-410. [doi: 10.6004/jnccn.2020.7631] [Medline: 33636694]
46. van den Boorn HG, Abu-Hanna A, Ter Veer E, van Kleef JJ, Lordick F, Stahl M, et al. SOURCE: A registry-based prediction model for overall survival in patients with metastatic oesophageal or gastric cancer. Cancers (Basel) 2019 Feb 05;11(2):187 [FREE Full text] [doi: 10.3390/cancers11020187] [Medline: 30764578]
47. Sivendran S, Latif A, McBride RB, Stensland KD, Wisnivesky J, Haines L, et al. Adverse event reporting in cancer clinical trial publications. J Clin Oncol 2014 Jan 10;32(2):83-89. [doi: 10.1200/JCO.2013.52.2219] [Medline: 24323037]
48. van de Water LF, van Kleef JJ, Dijksterhuis WP, Henselmans I, van den Boorn HG, Morel NM, et al. Communicating treatment risks and benefits to cancer patients: a systematic review of communication methods. Qual Life Res 2020 Jul;29(7):1747-1766 [FREE Full text] [doi: 10.1007/s11136-020-02503-8] [Medline: 32333238]
49. Nielsen J. Enhancing the explanatory power of usability heuristics. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 1994 Presented at: CHI94: ACM Conference on Human Factors in Computer Systems; April 24 - 28, 1994; Boston Massachusetts USA p. 152-158.
50. Sommerville I. Software Engineering 10th Edition (International Computer Science). Essex, UK: Pearson Education; 2016:1-808.
51. Hakone A, Harrison L, Ottley A, Winters N, Gutheil C, Han P, et al. PROACT: Iterative design of a patient-centered visualization for effective prostate cancer health risk communication. IEEE Trans Vis Comput Graph 2017 Jan;23(1):601-610. [doi: 10.1109/TVCG.2016.2598588] [Medline: 27875175]
52. Brooke J. SUS: a quick and dirty usability scale. In: Usability Evaluation In Industry. Boca Raton, Florida, United States: CRC Press; 1996.
53. Snijders T, Bosker R. Case Method Fast Track: A Rad Approach. Boston, Massachusetts, United States: Addison Wesley; 1994:1-224.

54. Villanueva R, Chen Z, Wickham H. ggplot2 : Elegant Graphics for Data Analysis. New York: Springer-Verlag; 2016:1-213.
55. Chang W, Cheng J, Allaire J, Xie Y, McPherson J. Shiny: web application framework for R. R Package Version 1.6.0. 2017. URL: https://cran.r-project.org/web/packages/shiny/index.html [accessed 2021-07-13]
56. Onwuegbuzie A, Dickinson W, Leech N, Zoran A. A qualitative framework for collecting and analyzing data in focus group research. Int J Qual Methods 2009 Sep 01;8(3):1-21. [doi: 10.1177/160940690900800301]
57. Think-aloud demonstration video. Gabor Aranyi PvSaPBoTU, UK (techslingtv). 2012. URL: https://www.youtube.com/watch?v=BwpPIiBK0cA [accessed 2021-02-12]
58. Kirschner P, van Merriënboer JJ. Ten steps to complex learning : a new approach to instruction and instructional design. Citeseer. URL: http://web.mit.edu/xtalks/TenStepsToComplexLearning-Kirschner-VanMerrienboer.pdf [accessed 2021-07-13]
59. Van Merrienboer JJ. Alternate models of instructional design: holistic design approaches and complex learning. In: Reiser R, Dempsey JV, editors. Trends and Issues in Instructional Design and Technology (2nd Ed.). Hoboken, New Jersey, United States: Merrill-Prentice Hall; 2007:72-81.
60. Van Merriënboer JJ, Kirschner PA. Three worlds of instructional design: state of the art and future directions. Instr Sci. 2001. URL: https://dspace.library.uu.nl/bitstream/handle/1874/18840/kirschner_01_three_worlds_instructional_design_state_art.pdf?sequence=1 [accessed 2021-07-14]
61. Henselmans I, Smets EM, de Haes JC, Dijkgraaf MG, de Vos FY, van Laarhoven HW. A randomized controlled trial of a skills training for oncologists and a communication aid for patients to stimulate shared decision making about palliative systemic treatment (CHOICE): study protocol. BMC Cancer 2018 Jan 08;18(1):55. [doi: 10.1186/s12885-017-3838-8] [Medline: 29310605]
62. Gilligan T, Coyle N, Frankel RM, Berry DL, Bohlke K, Epstein RM, et al. Patient-clinician communication: American Society of Clinical Oncology consensus guideline. J Clin Oncol 2017 Nov 01;35(31):3618-3632. [doi: 10.1200/JCO.2017.75.2311] [Medline: 28892432]
63. Kiely BE, Soon YY, Tattersall MH, Stockler MR. How long have I got? Estimating typical, best-case, and worst-case scenarios for patients starting first-line chemotherapy for metastatic breast cancer: a systematic review of recent randomized trials. J Clin Oncol 2011 Feb 01;29(4):456-463. [doi: 10.1200/JCO.2010.30.2174] [Medline: 21189397]
64. Kiely BE, Stockler MR, Tattersall MH. Thinking and talking about life expectancy in incurable cancer. Semin Oncol 2011 Jun;38(3):380-385. [doi: 10.1053/j.seminoncol.2011.03.007] [Medline: 21600367]
65. Zipkin DA, Umscheid CA, Keating NL, Allen E, Aung K, Beyth R, et al. Evidence-based risk communication: a systematic review. Ann Intern Med 2014 Aug 19;161(4):270-280. [doi: 10.7326/M14-0295] [Medline: 25133362]
66. Sella T, Botser D, Navon R, Biran H, Tenenbaum S, Urban D, et al. Preferences for disclosure of disease related information among thoracic cancer patients. Lung Cancer 2015 Apr;88(1):100-103. [doi: 10.1016/j.lungcan.2015.01.021] [Medline: 25682317]
67. Hoffmann TC, Del Mar C. Patients' expectations of the benefits and harms of treatments, screening, and tests: a systematic review. JAMA Intern Med 2015 Feb;175(2):274-286. [doi: 10.1001/jamainternmed.2014.6016] [Medline: 25531451]
68. Henselmans I, van Laarhoven HW, de Haes HC, Tokat M, Engelhardt EG, van Maarschalkerweerd PE, et al. Training for medical oncologists on shared decision-making about palliative chemotherapy: a randomized controlled trial. Oncologist 2019 Feb;24(2):259-265 [FREE Full text] [doi: 10.1634/theoncologist.2018-0090] [Medline: 29959285]
69. Colevas A, Setser A. The NCI Common Terminology Criteria for Adverse Events (CTCAE) v 3.0 is the new standard for oncology clinical trials. J Clin Oncol 2004 Jul 15;22(14_suppl):6098. [doi: 10.1200/jco.2004.22.90140.6098]
70. Scott N, Fayers P, Aaronson N, Bottomley A, de Graef A, Groenvold M, et al. EORTC QLQ-C30 reference values manual. EORTC Quality of Life Group. 2008. URL: https://www.eortc.org/app/uploads/sites/2/2018/02/reference_values_manual2008.pdf [accessed 2021-07-14]
71. Stiefel F, Barth J, Bensing J, Fallowfield L, Jost L, Razavi D, et al. Communication skills training in oncology: a position paper based on a consensus meeting among European experts in 2009. Ann Oncol 2010 Feb;21(2):204-207 [FREE Full text] [doi: 10.1093/annonc/mdp564] [Medline: 20026475]
72. De Vries AM, de Roten Y, Meystre C, Passchier J, Despland J, Stiefel F. Clinician characteristics, communication, and patient outcome in oncology: a systematic review. Psychooncology 2014 Apr;23(4):375-381. [doi: 10.1002/pon.3445] [Medline: 24243790]
73. Grossman R, Salas E. The transfer of training: what really matters. Int J Train Dev 2011;15(2):103-120 [FREE Full text] [doi: 10.1111/j.1468-2419.2011.00373.x]
74. Cepeda NJ, Coburn N, Rohrer D, Wixted JT, Mozer MC, Pashler H. Optimizing distributed practice: theoretical analysis and practical implications. Exp Psychol 2009;56(4):236-246. [doi: 10.1027/1618-3169.56.4.236] [Medline: 19439395]
75. Dempster F. Spacing effects and their implications for theory and practice. Educ Psychol Rev 1989 Dec;1(4):309-330. [doi: 10.1007/BF01320097]
76. Dempster F, Perkins P. Revitalizing classroom assessment: using tests to promote learning. J Instr Psychol 1993;20(3):197-203 [FREE Full text]
77. Donovan J, Radosevich D. A meta-analytic review of the distribution of practice effect: now you see it, now you don't. J Appl Psychol 1999;84(5):795-805. [doi: 10.1037/0021-9010.84.5.795]

78. Moss V. The efficacy of massed versus distributed practice as a function of desired learning outcomes and grade level of the student. All Graduate Theses and Dissertations - Utah State University. 1995. URL: https://digitalcommons.usu.edu/about.html [accessed 2021-07-14]
79. Janiszewski C, Noel H, Sawyer A. A meta-analysis of the spacing effect in verbal learning: implications for research on advertising repetition and consumer memory. J Consum Res 2003 Jun 01;30(1):138-149. [doi: 10.1086/374692]
80. van Kleef JJ, van den Boorn HG, Verhoeven R, Vanschoenbeek K, Abu-Hanna A, Zwinderman A, et al. External validation of the Dutch Source Survival Prediction Model in Belgian metastatic oesophageal and gastric cancer patients. Cancers (Basel) 2020 Mar 31;12(4):834 [FREE Full text] [doi: 10.3390/cancers12040834] [Medline: 32244310]
81. Hillen MA, Gutheil CM, Smets EM, Hansen M, Kungel TM, Strout TD, et al. The evolution of uncertainty in second opinions about prostate cancer treatment. Health Expect 2017 Dec;20(6):1264-1274 [FREE Full text] [doi: 10.1111/hex.12566] [Medline: 28521078]
82. van den Boorn HG, Stroes CI, Zwinderman AH, Eshuis WJ, Hulshof MC, van Etten-Jamaludin FS, et al. Health-related quality of life in curatively-treated patients with esophageal or gastric cancer: a systematic review and meta-analysis. Crit Rev Oncol Hematol 2020 Oct;154:103069 [FREE Full text] [doi: 10.1016/j.critrevonc.2020.103069] [Medline: 32818901]
83. van den Ende T, Nijenhuis FA, van den Boorn HG, Ter Veer E, Hulshof MC, Gisbertz SS, et al. COMplot, a graphical presentation of complication profiles and adverse effects for the curative treatment of gastric cancer: a systematic review and meta-analysis. Front Oncol 2019;9:684 [FREE Full text] [doi: 10.3389/fonc.2019.00684] [Medline: 31403035]
84. Source Tool. 2019. URL: https://source.amc.nl/ [accessed 2021-02-12]

Abbreviations

CHOICE: Choosing Treatment Together in Cancer at the End of Life
EORTC QLQ-C30: European Organisation for Research and Treatment of Cancer quality of life questionnaire C30
HRQoL: health-related quality of life
MoSCoW: Must Have, Should Have, Could Have, Won't Have
SUS: System Usability Scale
TNM: tumor–node–metastasis

Edited by R Kukafka; submitted 12.02.21; peer-reviewed by A Freeman, A Naser, A Kanaan; comments to author 12.03.21; revised version received 07.05.21; accepted 24.05.21; published 27.08.21.

Please cite as:
van de Water LF, van den Boorn HG, Hoxha F, Henselmans I, Calff MM, Sprangers MAG, Abu-Hanna A, Smets EMA, van Laarhoven HWM
Informing Patients With Esophagogastric Cancer About Treatment Outcomes by Using a Web-Based Tool and Training: Development and Evaluation Study
J Med Internet Res 2021;23(8):e27824
URL: https://www.jmir.org/2021/8/e27824
doi:10.2196/27824
PMID:34448703

©Loïs F van de Water, Héctor G van den Boorn, Florian Hoxha, Inge Henselmans, Mart M Calff, Mirjam A G Sprangers, Ameen Abu-Hanna, Ellen M A Smets, Hanneke W M van Laarhoven. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Effects of the Interactive Web-Based Video “Mon Coeur, Mon BASIC” on Drug Adherence of Patients With Myocardial Infarction: Randomized Controlled Trial

Christel Bruggmann[1,2,3], PharmD, PhD; Julien Adjedj[4], MD, PhD; Sylvain Sardy[5], PhD, Prof Dr; Olivier Muller[4], MD, Prof Dr; Pierre Voirol[1,2], PhD; Farshid Sadeghipour[1,2], PhD, Prof Dr

[1]Department of Pharmacy, University Hospital of Lausanne, University of Lausanne, Lausanne, Switzerland
[2]Institute of Pharmaceutical Sciences of Western Switzerland, University of Geneva, Geneva, Switzerland
[3]Department of Pharmacy, University Hospital of Geneva, Geneva, Switzerland
[4]Department of Cardiology, University Hospital of Lausanne, University of Lausanne, Lausanne, Switzerland
[5]Section of Mathematics, University of Geneva, Geneva, Switzerland

Corresponding Author:
Christel Bruggmann, PharmD, PhD
Department of Pharmacy
University Hospital of Geneva
Rue Gabrielle Perret-Gentil 4
Geneva, 1205
Switzerland
Phone: 41 796146714
Email: cbruggmann@gmail.com

Abstract

Background: Secondary prevention strategies after acute coronary syndrome (ACS) presentation with the use of drug combinations are essential to reduce the recurrence of cardiovascular events. However, lack of drug adherence is known to be common in this population and to be related to treatment failure. To improve drug adherence, we developed the “Mon Coeur, Mon BASIC” video. This online video has been specifically designed to inform patients about their disease and their current medications. Interactivity has been used to increase patient attention, and the video can also be viewed on smartphones and tablets.

Objective: The objective of this study was to assess the long-term impact of an informative web-based video on drug adherence in patients admitted for an ACS.

Methods: This randomized study was conducted with consecutive patients admitted to University Hospital of Lausanne for ACS. We randomized patients to an intervention group, which had access to the web-based video and a short interview with the pharmacist, and a control group receiving usual care. The primary outcome was the difference in drug adherence, assessed with the Adherence to Refills and Medication Scale (ARMS; 9 multiple-choice questions, scores ranging from 12 for perfect adherence to 48 for lack of adherence), between groups at 1, 3, and 6 months. We assessed the difference in ARMS score between both groups with the Wilcoxon rank sum test. Secondary outcomes were differences in knowledge, readmissions, and emergency room visits between groups and patients’ satisfaction with the video.

Results: Sixty patients were included at baseline. The median age of the participants was 59 years (IQR 49-69), and 85% (51/60) were male. At 1 month, 51 patients participated in the follow-up, 50 patients participated at 3 months, and 47 patients participated at 6 months. The mean ARMS scores at 1 and 6 months did not differ between the intervention and control groups (13.24 vs 13.15, 13.52 vs 13.68, respectively). At 3 months, this score was significantly lower in the intervention group than in the control group (12.54 vs 13.75; P=.03). We observed significant increases in knowledge from baseline to 1 and 3 months, but not to 6 months, in the intervention group. Readmissions and emergency room visits have been very rare, and the proportion was not different among groups. Patients in the intervention group were highly satisfied with the video.

Conclusions: Despite a lower sample size than we expected to reach, we observed that the “Mon Coeur, Mon BASIC” web-based interactive video improved patients’ knowledge and seemed to have an impact on drug adherence. These results are encouraging, and the video will be offered to all patients admitted to our hospital with ACS.

Trial Registration: ClinicalTrials.gov NCT03949608; https://clinicaltrials.gov/ct2/show/NCT03949608

(J Med Internet Res 2021;23(8):e21938) doi:10.2196/21938

KEYWORDS

acute coronary syndrome; eHealth; drug adherence; mHealth; mobile phone

Introduction

Cardiovascular disease is a major cause of morbi-mortality in Europe, with a substantial contribution from acute coronary syndrome (ACS) [1]. Secondary prevention strategies, such as risk factor control through lifestyle modifications and the use of medication combinations, have greatly reduced the recurrence of ACS [2]. European guidelines have been developed to enhance evidence-based medicine prescriptions for patients with myocardial infarction (MI) [3,4], and we previously showed that physicians at our hospital in Switzerland predominantly issued prescriptions of this type [5]. However, poor patient self-adherence to cardiac medications has been documented worldwide [6-8] and has been associated with increased morbi-mortality [9-11]. The discontinuation of antiplatelet drugs has been related to fatal consequences such as stent thrombosis, particularly soon after ACS onset [12].

Poor drug adherence is related to many factors associated with health care systems (eg, cost, access to care), socioeconomics (ie, poverty), therapy (ie, treatment complexity, cost), and patients (ie, health literacy, willingness to change, knowledge, education) [13]. In Switzerland, the health care system offers high-quality care to all residents, and less social inequality exists than in other countries; we thus believe that patient adherence in this country is more likely related to treatment and patient factors. The provision of sufficient and effective information to patients with chronic diseases has been shown to increase patient satisfaction [14], reduce psychological distress [15], enhance patients' perceived control [16], and improve patient adherence to medication prescriptions [13]. Thus, the offering of such information and knowledge to all patients with ACS is very important.

Various interventions have been shown to promote drug adherence in the context of cardiovascular health. These strategies range from the simple, such as the distribution of written material about medications [17] and the regular mailing of informational letters [18,19], to multifaceted interventions (eg, medication reconciliation, therapeutic education) involving clinical pharmacists [20,21] and nurses specialized in therapeutic education [22].

The length of hospital stays after MI has shortened in recent years, which has reduced the number of opportunities to inform patients about their disease during hospitalization. Cardio-rehabilitation (CR) centers are meant to fill this gap, but not all patients participate in CR programs. In addition, depending on the type of program chosen (ie, stationary or ambulatory), the objective may be oriented more toward cardiovascular reeducation than to patient education. Moreover, very few sessions are devoted to educating patients about their drug treatments.

Different kinds of mobile health (mHealth) technologies have been developed to support cardiovascular health, like eHealth diaries [23], apps supporting cardiac rehabilitation [24,25], and complete e-learning platforms [26]. However, none of these technologies were available in French, and only a few studies have evaluated the impact of mHealth technologies on therapeutic adherence [27,28]. As we were convinced that mHealth technologies have an impact on drug adherence, we wanted to test it with our study. Therefore, we developed a new approach using a video tool for the provision of information and patient education. The tool we developed is interactive, web-based, smartphone- and tablet-compatible, and it is designed to be offered to patients with ACS during their hospital stays. We tested whether this tool has impacts on drug adherence and patient knowledge after ACS. Our study is therefore expected to fill some important gaps in the current literature.

Methods

Study Design and Population

The Secondary Prevention of ACS With Beta-Blockers, Antiaggregants, Statins, Angiotensin-Converting Enzyme Inhibitors, and Risk Factor Control (BASIC) study was a single-center randomized trial. The objective of the study was to assess if a new web-based and interactive video could increase long-term drug adherence, as well as the knowledge of the patients about their disease and their current medications, in patients admitted in hospital for an ACS.

The patients were screened between February 1 and September 1, 2019, on admission to the University Hospital of Lausanne. The intervention for this study was added to usual care, in which all patients with ACS watched a 20-minute video called ELIPS during their hospital stays [29]. ELIPS explains acute infarction, acute care, and the drugs prescribed. Patients also received a booklet about coronary artery disease. All patients admitted to our hospital for ACS are encouraged to participate in a CR program after discharge; the program chosen depends on each patient's willingness and home location. Two types of programs are offered (a 3-week stationary program and an ambulatory program), and both are reimbursed by health insurance.

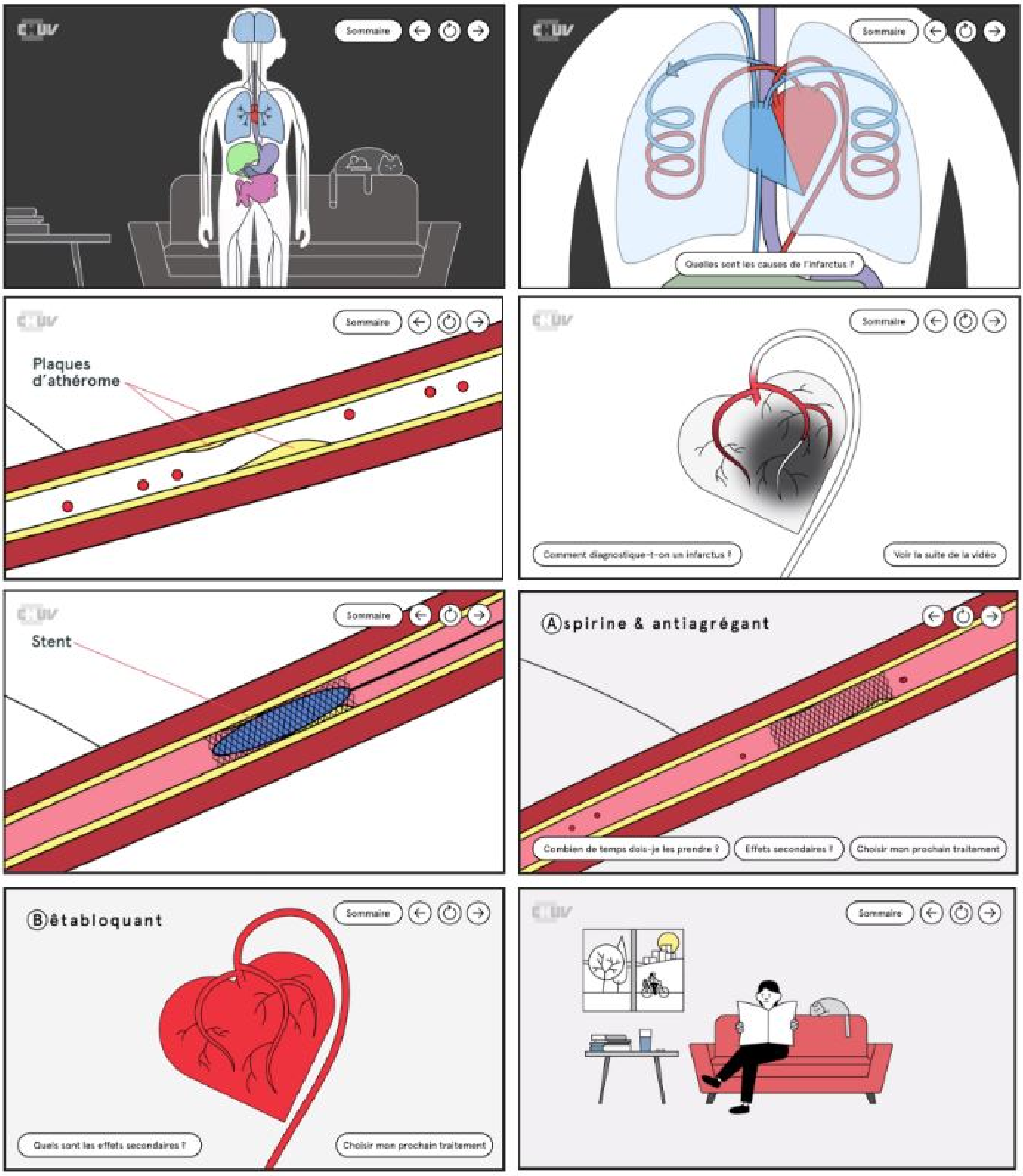
Participants eligible for the study were men and women older than 18 years who were diagnosed with ST-segment elevation myocardial infarction (STEMI) or non–ST-segment elevation myocardial infarction (NSTEMI) treated with percutaneous coronary intervention (PCI). Other requirements were total discernment capacity; possession of a digital tablet, smartphone, or home computer; and satisfactory French language skills. Exclusion criteria were the inability to follow the study procedure (eg, due to language problems, psychological disorders, dementia), refugee claimant status, homelessness or incarceration (because of the impossibility of contacting such individuals after discharge), and life expectancy <6 months due to another disease.

Eligible participants were randomized to the intervention and control groups using a 1:1 allocation ratio after providing oral and written informed consent. The randomization was conducted by a clinical study specialist who was not involved in the study. The randomization process was made by week and block of 4. Each week was randomized as intervention or control in order to avoid having two patients in different groups next to each other in the intermediate care unit.

"Mon Coeur, Mon BASIC" Web-Hosted Video

"Mon Coeur, Mon BASIC" is an interactive web-hosted video that is smartphone- and tablet-compatible. It is freely available online [30]. It is a simple animated cartoon (Figure 1) with narration in French. It consists of three parts providing information about heart function and the physiopathology of ACS, acute care for ACS (coronarography and PCI), and the medications prescribed after ACS (usefulness and side effects), with a total length of about 15 minutes. The viewer can click on the video using a mouse or finger to obtain information about a particular point or to jump directly to another subject. The video was developed by a working group consisting of a pharmacist, a graphic designer, a specialist in patient communication, and a cardiologist. It was designed using Illustrator and InDesign (both from Adobe Inc, San José, California), and the interactivity was created using Storyline 360 (Articulate, New York). It was tested with several patients, and all bugs were eliminated before the beginning of the study.

Figure 1. Screenshots from the “Mon Coeur, Mon BASIC” video.


CHUV
Sommaire
Quelles sont les causes de l'infarctus ?
Plaques d'athérome
Comment diagnostique-t-on un infarctus ?
Voir la suite de la vidéo
Stent
Aspirine & antiagrégant
Combien de temps dois-je les prendre ?
Effets secondaires ?
Choisir mon prochain traitement
Bêtabloquant
Quels sont les effets secondaires ?
Choisir mon prochain traitement

Intervention and Control Condition

The pharmacist participating in the study met all patients in the intervention group during their hospital stays and gave each of them a medication card with his or her current prescriptions (Multimedia Appendix 1). The medication card contained all the medications currently prescribed to the patient, classified by type of medication according to the acronym BASIC for Beta-blocker, Antiaggregant, Statin, angiotensin-converting enzyme Inhibitor, and Control of risk factors (ie, smoke cessation, limitation of fatty meals, etc), and was the size of a credit card. She then connected to the e-learning website on the patient’s tablet or smartphone and created a shortcut. Tablets from the institution were lent to patients with no tablet, smartphone, or computer. The pharmacist asked the patients to watch the video at their convenience. After the patients had watched the video, the pharmacist visited to answer questions and distribute study questionnaires (described below). The postvideo visit was short, and the aim of the visit was to check if the patients had watched the video and to answer patients’ questions if they had any. The duration of this visit was around 10 minutes per patient.

The control group received usual care. In usual care, the patient has no meeting with the pharmacist at all, and the information about medications and the disease are provided by the nurses and the physician in charge of the patient during hospitalization. All the patients (in control and intervention group) also watched a short film developed in the Geneva University Hospitals [29]. For the purpose of the study, the pharmacist only met patients randomized to the control group to give them the study questionnaires.

Primary Outcome

The primary outcome was the difference in medication adherence between groups at 1, 3, and 6 months after ACS for the assessment of postdischarge treatment initiation, treatment implementation in daily life, and long-term treatment persistence, respectively. Adherence was assessed using the Adherence to Refills and Medication Scale (ARMS; Multimedia Appendix 2). This self-report questionnaire was chosen due to its strong internal consistency (Cronbach α=.814), good correlation with other subjective and objective measures, and validation for use with patients with coronary heart disease [32]. It consists of 12 items assessing adherence to taking medication (n=8) and refilling prescriptions (n=4). The questionnaire is made up of affirmations that can be answered with "none of the time," "some of the time," "most of the time," or "all of the time," varying from 1 to 4 points, respectively. Final scores range from 12 (most adherence) to 48 (least adherence) and can be treated as continuous measures or dichotomized as 12 and >12. We translated the validated English version of the ARMS into French according to guidelines developed for health care research [33].

Secondary Outcomes

The first secondary outcome was the difference in ACS-related knowledge (basics of heart function, ACS pathophysiology, and usefulness of prescribed medications) between groups, assessed using a questionnaire developed specifically for this study (Multimedia Appendix 3). The questionnaire has 9 items, and scores range from 0 to 9 (most knowledge). The questionnaire was given to the intervention group as a pretest (before the video viewing) and posttest (after viewing), to the control group 1 day after study inclusion, and to all participants at 1, 3, and 6 months. We also assessed whether the groups differed in a composite measure of cardiovascular mortality, first occurrence of reinfarction, recurrence of ACS, cardiovascular death, and readmission and emergency room visitation over the 6-month study period. Finally, we assessed patients' satisfaction with the video using a questionnaire of 11 multiple-choice questions.

Data Collection

Most baseline data (eg, demographic characteristics, laboratory values, and vital parameters on day 2 after ACS; ACS type and therapeutic strategy; cardiovascular risk factors; drug prescribed at discharge) were collected from computerized patient records generated during hospitalization. Other data were collected through patient interviews; they included patients' email addresses, educational levels (graduation of primary, secondary, or tertiary school), employment statuses (full time, part time, retired, or unemployed), spoken French levels (native, near native, highly proficient, very good working knowledge, or basic communication skills), general practitioners' names, types of device used at home (smartphone, tablet, or computer), information and communications technology (ICT) use levels (low: short message service or telephone only, medium: also maps and basic online research, or high: many applications in daily life), and health literacy scores, assessed with the validated French translation of the Functional, Communicative and Critical Health Literacy (FCCHL) tool [34]. FCCHL scores range from 14 (least literacy) to 70 (most literacy). All data were collected and managed using REDCap (Research Electronic Data Capture) tools hosted in the University Hospital of Lausanne [35,36]. REDCap is a secure, web-based software platform designed to support data capture for research studies.

An email with a link to the study questionnaires (hosted by REDCap) was sent to the participants at 1, 3, and 6 months after study inclusion. For participants who did not provide any email address, the questionnaires were mailed with postage-paid return envelopes.

Statistical Analysis

An intention-to-treat analysis including all study participants was performed. The characteristics of patients randomized to the intervention and control groups were expressed as medians with 25th and 75th percentiles for continuous variables, and as numbers and percentages for categorical variables. They were compared using the Pearson chi-square test for categorical variables and the *t* test for continuous variables.

The ARMS scores at 1, 3, and 6 months were compared between groups using the Wilcoxon rank sum test. This test was also used to compare pretest and posttest knowledge scores and those at 1, 3, and 6 months between groups. The Wilcoxon rank sum test for paired data was used to assess differences in knowledge within groups between study timepoints. Descriptive statistics (numbers and percentages) were calculated for the composite outcome, satisfaction scores, and video reuse. The level of significance for all analyses was set at two-sided α<.05. All analyses were performed using Stata software (version 14; StataCorp, College Station, Texas).

We performed a power analysis to estimate the sample size required to detect a significant difference in medication adherence between the intervention and control groups. We based the calculation on the mean ARMS score of 16.32 (SD 4.06) for chronically ill patients [32]. We determined that a sample of 128 patients (64 per group) was needed to detect a difference of 0.5 SD (ie, 2.03) between the intervention and control groups at a two-sided 5% significance level with a power of 0.8 and an allocation ratio of 1:1. Assuming 10% loss to follow-up, the target sample size was 142 patients. Based on our hospital's annual admission rate of 500 patients with NSTEMI and 300 patients with STEMI, we initially believed that study enrollment would be completed within 20 weeks, but because many patients were discharged before we could conduct baseline assessment, and an unexpectedly large number of patients refused to participate, we extended the enrollment period to 37 weeks. However, because the investigator was not available for a longer period of time than planned, we had to terminate the study before the target sample size had been reached. The results presented in this article must therefore be interpreted with caution because the study was underpowered.

Ethical Considerations

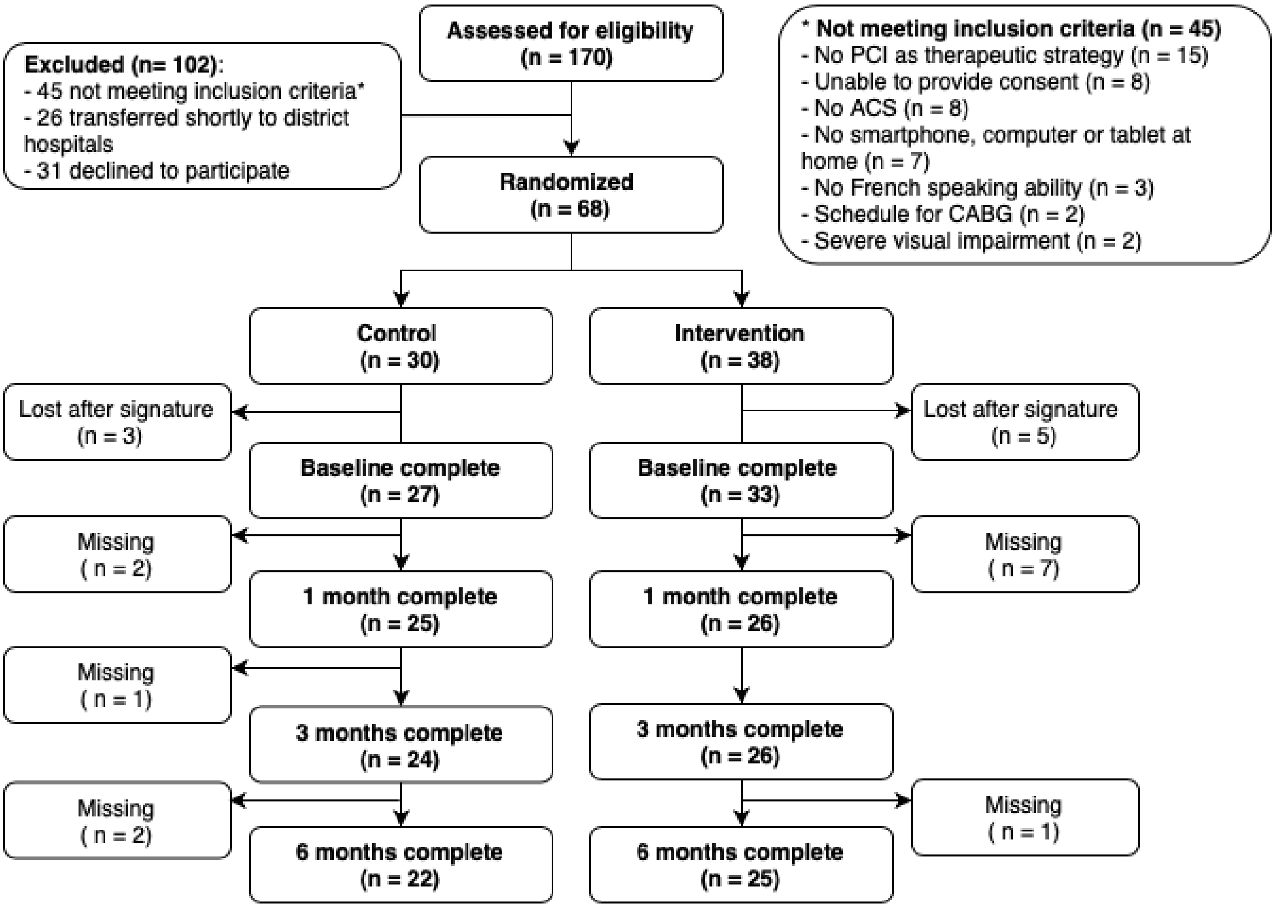
This study was approved by the local ethics committee (*Commission cantonale d'éthique de la Recherche sur l'être humain du Canton de Vaud*; number 2018-02223) and registered at ClinicalTrials.gov (number NCT03949608). It complied with the principles of good clinical practice and the Declaration of Helsinki.

Results

Study Population and Baseline Characteristics

In total, 170 patients were screened during the study period, and 125 patients were asked to participate (Figure 2). Of these, 31 (24.8%) patients declined participation, and 26 (20.8%) were transferred to district hospitals shortly thereafter and were not enrolled. We randomized 68 patients to the control (n=30) and intervention (n=38) groups. Of these, 8 patients were transferred unexpectedly to district hospitals, did not watch the video, or did not complete the baseline questionnaires; thus, the final baseline sample consisted of 60 patients (33 in the intervention group and 27 in the control group) with complete baseline data. An additional 21 of the 68 randomized patients (31%) did not complete the follow-up questionnaires.

Figure 2. Patient flowchart within the study period. ACS: acute coronary syndrome; CABG: coronary artery bypass grafting; PCI: percutaneous coronary intervention.


Assessed for eligibility (n = 170)
Excluded (n= 102):
- 45 not meeting inclusion criteria*
- 26 transferred shortly to district hospitals
- 31 declined to participate
* Not meeting inclusion criteria (n = 45)
- No PCI as therapeutic strategy (n = 15)
- Unable to provide consent (n = 8)
- No ACS (n = 8)
- No smartphone, computer or tablet at home (n = 7)
- No French speaking ability (n = 3)
- Schedule for CABG (n = 2)
- Severe visual impairment (n = 2)
Randomized (n = 68)
Control (n = 30)
Intervention (n = 38)
Lost after signature (n = 3)
Lost after signature (n = 5)
Baseline complete (n = 27)
Baseline complete (n = 33)
Missing (n = 2)
Missing (n = 7)
1 month complete (n = 25)
1 month complete (n = 26)
Missing (n = 1)
3 months complete (n = 24)
3 months complete (n = 26)
Missing (n = 2)
Missing (n = 1)
6 months complete (n = 22)
6 months complete (n = 25)

The median age of the 60 patients included was 59 years; 85% (51/60) of the patients were men, 19% (11/59) had diabetes, and about 45% (25/56) were active smokers (Table 1). Patients in the control group were slightly older and more often retired than patients in the intervention group were. They also had more comorbidities and more often had NSTEMI than did patients in the intervention group. However, the groups were well matched, with no significant difference in baseline characteristics.

Table 1. Baseline patient characteristics (N=60). Categorical data are presented as n (%) and continuous data are presented as median (interquartile range). *P* values were estimated with the chi-square test for categorical variables and *t* test for continuous variables.

Characteristic	Overall (N=60)	Intervention (n=33)	Control (n=27)	*P* value
Demographic characteristics				
Sex (male), n (%)	51 (85)	27 (82)	24 (89)	.45
Age (years), median (IQR)	59 (49-69)	56 (46-68)	62 (56-71)	.10
Civil status, n (%)				.70
Single	10 (17)	4 (12)	6 (22)	
Married	36 (60)	20 (61)	16 (59)	
Divorced	11 (18)	7 (21)	4 (15)	
Widowed	3 (5)	2 (6)	1 (4)	
Education level, n (%)				.54
Compulsory school	8 (13)	6 (18)	2 (7)	
High school, internship	28 (47)	15 (45)	13 (48)	
University of applied sciences	8 (13)	3 (9)	5 (18.5)	
Bachelor’s degree	5 (8)	2 (6)	3 (11.1)	
Master’s degree or more	11 (18)	7 (21)	4 (14.8)	
Employment, n (%)				.17
Full time	29 (48)	20 (61)	9 (33.3)	
Part time	7 (12)	2 (6)	5 (19)	
Retired	20 (33)	9 (27)	11 (41)	
Unemployed	4 (7)	2 (6)	2 (7)	
French speaking level, n (%)				.66
Native	48 (80)	26 (79)	22 (81)	
Near native	12 (20)	7 (21)	5 (19)	
Device type, n (%)				
Computer	55 (92)	30 (91)	25 (93)	.81
Smartphone	51 (85)	28 (85)	23 (85)	
Tablet	38 (63)	24 (73)	14 (52)	
ICT[a] use level, n (%)				.37
Low	12 (20)	7 (21)	5 (19)	
Medium	9 (15)	3 (9)	6 (22)	
High	39 (65)	23 (70)	16 (59)	
Health literacy (FCCHL[b]) score, median (IQR)	49 (44.5-54)	50 (44-54)	48 (45-54)	.95
Cardiovascular risk factors				
Smoking status[c], n (%)				.13
Active	25 (45)	14 (47)	11 (42)	
Former	17 (30)	9 (30)	8 (31)	
Diabetes[d], n (%)	11 (19)	5 (16)	6 (22)	.52
Hypertension, n (%)	32 (53)	17 (52)	15 (56)	.76
Dyslipidemia[e], n (%)	34 (59)	23 (72)	11 (42)	.02
Alcohol consumption[f], n (%)	5 (10)	3 (9)	2 (11)	.89
Family history of ACS[g,h], n (%)	24 (44)	15 (47)	9 (41)	.67

Characteristic	Overall (N=60)	Intervention (n=33)	Control (n=27)	*P* value
Overweight, n (%)	37 (62)	21 (64)	16 (59)	.73
Drugs used chronically before ACS, median (IQR)	1 (0-3)	0 (0-2)	2 (1-3)	.12
ACS type, n (%)				
STEMI[i]	52 (87)	31 (94)	21 (78)	.07
NSTEMI[j]	8 (13)	2 (6)	6 (22)	
LVEF[k] evaluation during hospitalization, n (%)				.23
<40%	6 (10)	2 (6)	4 (15)	
≥40%	54 (90)	31 (94)	23 (85)	
Discharge				
Total length of stay (days), median (IQR)	3 (2-4)	3 (2-4)	3 (2-4)	.70
Destination after university hospital discharge, n (%)				.71
Home	40 (67)	23 (70)	17 (63)	
Cardio-rehabilitation center	2 (3)	1 (3)	1 (4)	
District hospital	17 (28)	8 (24)	9 (33)	
Another in-hospital department	1 (2)	1 (3)	0 (0)	
Drugs prescribed at discharge, median (IQR)	6 (6-7.5)	6 (6-7)	7 (5-9)	.53
Prescriptions at discharge, n (%)				
Beta-blocker	53 (88)	27 (82)	26 (96)	.08
ACEI[l]	55 (92)	30 (91)	25 (93)	.81
Statin	59 (98)	33 (100)	26 (96)	.27
Aspirin	60 (100)	33 (100)	27 (100)	N/A[m]
$P2Y_{12}$ inhibitor	60 (100)	33 (100)	27 (100)	N/A

[a] ICT: information and communications technology.

[b] FCCHL: Functional, Communicative and Critical Health Literacy tool (French translation).

[c] Four missing values (3 in the intervention group, 1 in the control group).

[d] One missing value (intervention group).

[e] Two missing values (1 in each group).

[f] Nine missing values (1 in the intervention group, 8 in the control group).

[g] Six missing values (1 in the intervention group, 5 in the control group).

[h] ACS: acute coronary syndrome.

[i] STEMI: ST-segment elevation myocardial infarction.

[j] NSTEMI: non–ST-segment elevation myocardial infarction.

[k] LVEF: left ventricular ejection fraction.

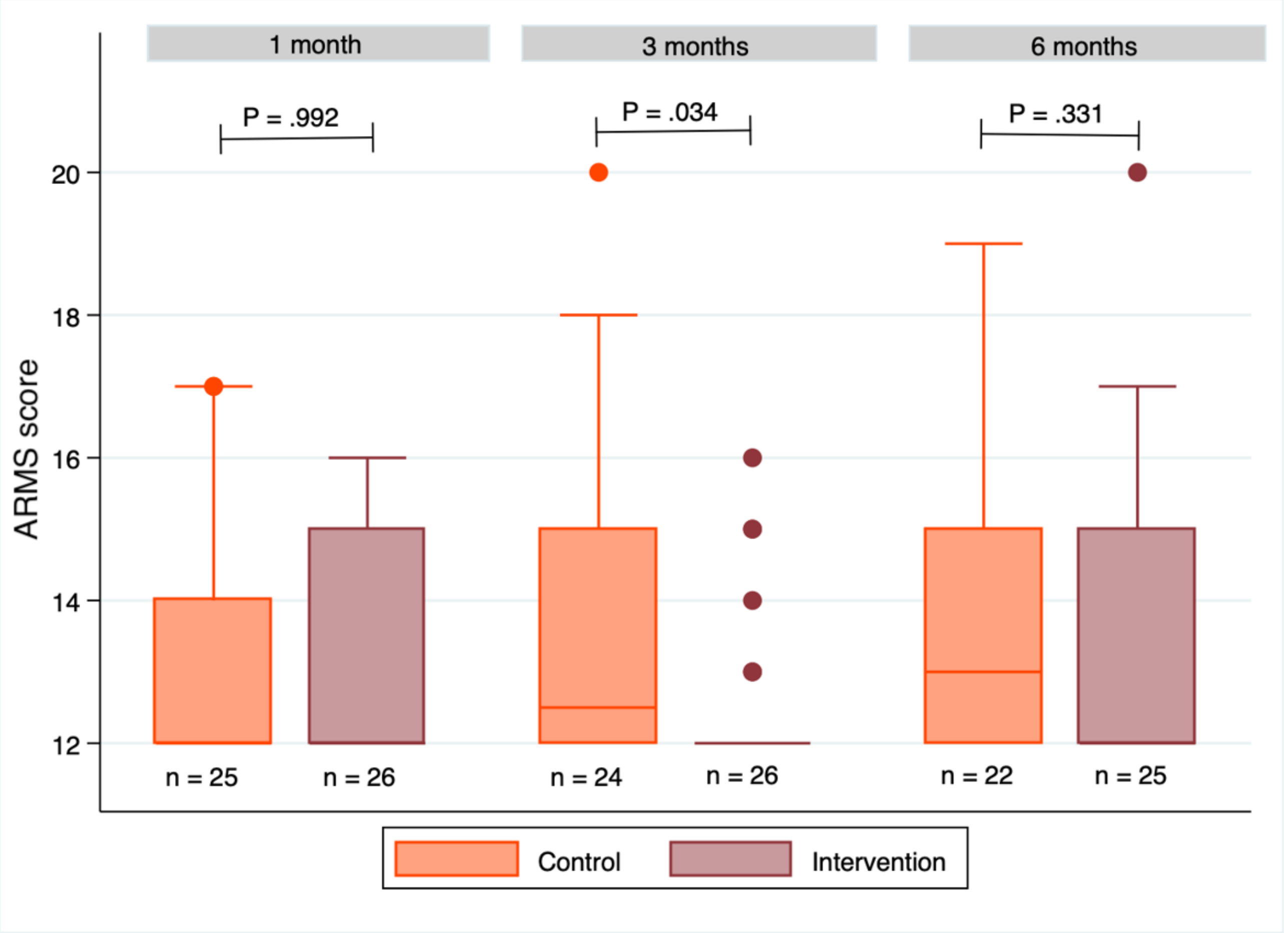
[l] ACEI: angiotensin-converting enzyme inhibitor.

[m] N/A: not applicable.

Primary Outcome: ARMS Score

At 1 month, the mean ARMS score did not differ significantly between the intervention and control groups (13.15, 95% CI 12.56-13.74 and 13.24, 95% CI 12.52-13.96, respectively; *P*=.99). At 3 months, the ARMS score was significantly lower in the intervention group than in the control group (12.54, 95% CI 12.08-13.00 vs 13.75, 95% CI 12.74-14.76; *P*=.03). At 6 months, this score did not differ between the intervention and control groups (13.52, 95% CI 12.63-14.41 and 13.68, 95% CI 12.96-14.76, respectively; *P*=.33; Figure 3). The median ARMS score increased from 1 to 6 months in the control group but remained more stable (at ~12) in the intervention group, despite a change in distribution.

Figure 3. ARMS scores at 1, 3, and 6 months. In the box plots, the boundary of the box closest to zero indicates the 25th percentile, the line within the box represents the median, and the boundary of the box farthest from zero indicates the 75th percentile. Whiskers above the box indicate the 90th percentile. Points above the whiskers represent the outliers outside the 90th percentile. The *P* values represent statistics made with the nonparametric Wilcoxon test. n represents the number of participants per group. ARMS: Adherence to Refills and Medication Scale.



Of 13 participants whose ARMS scores increased between 1 and 3 months, 10 were in the intervention group, and 3 were in the control group. In 7 (54%) cases, scores increased because patients provided different responses to the last item about prescription refills (Multimedia Appendix 2).

Secondary Outcomes

Knowledge Score

The mean knowledge score did not differ significantly between the intervention and control groups at baseline or at 1, 3, or 6 months (7.22, 95% CI 6.64-7.81 and 7.03, 95% CI 6.46-7.60; 8.19, 95% CI 7.66-8.72 and 7.72, 95% CI 7.06-8.38; 8.36, 95% CI 8.03-8.69 and 8.00, 95% CI 7.44-8.56; and 8.04, 95% CI 7.54-8.54 and 7.72, 95% CI 7.00-8.45, respectively; all *P*s>.05; Figure 4). Within each group, knowledge increased from baseline to 6 months; increases were significant in the intervention group between baseline and 3 months, but not at 6 months (Table 2).

Figure 4. Knowledge scores (0-9) at baseline (pretest), 1 day after ACS (posttest), and 1, 3, and 6 months after ACS. In the box plots, the boundary of the box closest to zero indicates the 25th percentile, the line within the box represents the median, and the boundary of the box farthest from zero indicates the 75th percentile. Whiskers above the box indicate the 90th percentile. Points above the whiskers represent the outliers outside the 90th percentile. n represents the number of participants per group. ACS: acute coronary syndrome.

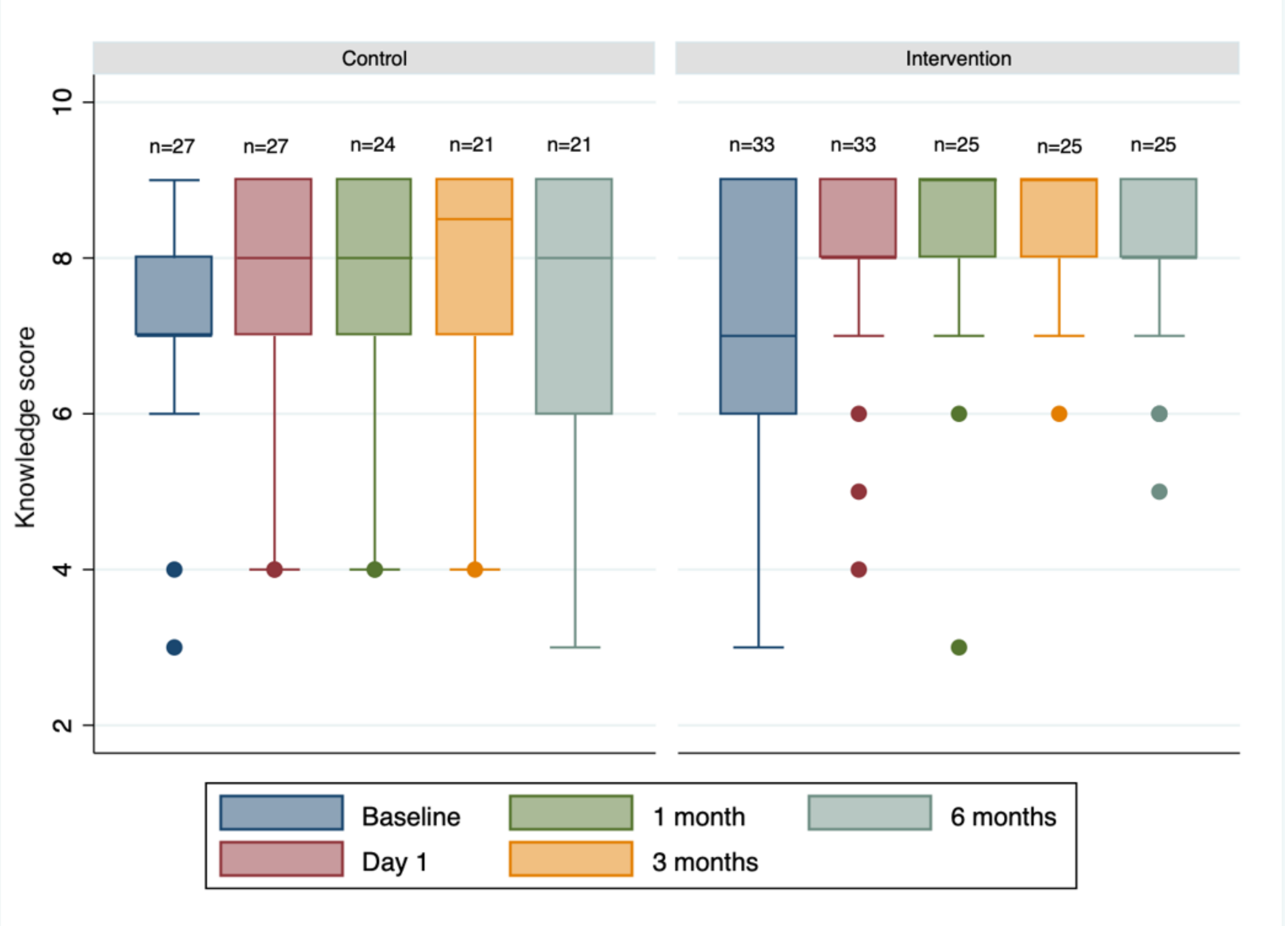


Table 2. Differences in knowledge scores from baseline.

Timepoint	Control, median (95% CI)	*P* value[a]	Intervention, median (95% CI)	*P* value[a]
Baseline score	7.22 (6.64 to 7.81)	N/A[b]	7.03 (6.46 to 7.60)	N/A
Difference from baseline				
Posttest (1 day)	0.33 (–0.16 to 0.82)	.14	1.00 (0.48 to 1.52)	.001
1 month	0.48 (–1.13 to 1.19)	.17	1.04 (0.35 to 1.73)	.03
3 months	0.63 (–0.02 to 1.27)	.06	1.40 (0.68 to 2.12)	.008
6 months	0.32 (–0.30 to 0.94)	.27	0.88 (0.25 to 1.51)	.144

[a] Wilcoxon rank sum test for paired data.

[b] N/A: not applicable.

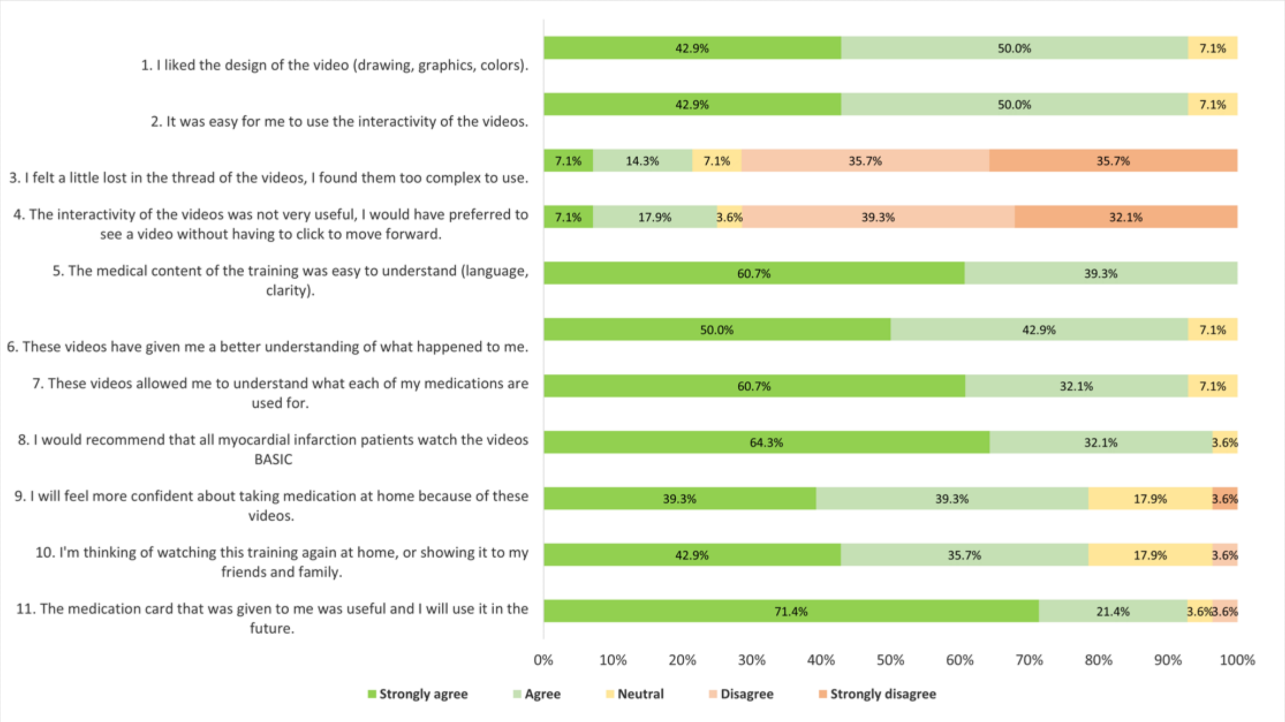
Composite Endpoint of Mortality, Reinfarction, Rehospitalization, and Emergency Room Visits

No death occurred in the cohort. Five patients visited the emergency room and were subsequently hospitalized. Overall, 7 of 46 (15%) patients (2 in the intervention group, 5 in the control group) were hospitalized for cardiovascular reasons (heart rhythm disorders: 1 tachycardia, 1 bradycardia; vagal discomfort: n=2; chest pain that was determined to be noncardiac: n=1; elective stent placement: n=1; and coronary artery bypass grafting: n=1). No patient had a new infarct during the follow-up period and no difference was observed between groups.

Satisfaction With and Reuse of the Video and Medication Card

Overall, the patients were highly satisfied with the design, use, and medical content of the video (Figure 5). The majority of patients felt that the video helped them to better understand their disease and medications and indicated that they would recommend it to other patients with ACS. As questions 3 and 4 were negatively worded, some patients may have answered them incorrectly. Despite this potential issue, the majority of patients felt comfortable with the interactivity of the video and did not lose the thread of the presentation. Responses regarding whether patients thought they would watch the video again at home were the most mitigated. We found that 18 of 30 (60%) patients viewed the video again at home. The majority of patients strongly appreciated the medication cards, and 22 of 30 (73%) patients reused the cards after discharge.

Figure 5. Patients’ satisfaction with the video (n=28).



Discussion

Principal Findings

The “Mon Coeur, Mon BASIC” web-hosted video seemed to significantly improve the drug adherence of patients with ACS for a few months after treatment initiation: a significant effect was observed at 3 months, but not at 1 or 6 months. Moreover, patients from the intervention group showed significant improvement in their knowledge about ACS and their medications from baseline to 3 months; no such effect was seen in the control group. In addition, patients were generally very satisfied with the information provided in the video and with the medication cards given to them.

The rate of drug adherence was high in both groups at 1 month after ACS. As expected, the median ARMS score tended to increase over time in the control group, reflecting decreasing adherence. At 3 months, the intervention group showed much better drug adherence than did the control group. At 6 months, the median score remained lower in the intervention group, but this difference was not significant. Although the observed trends may not all be attributable to the effect of the video alone, we can state with strong confidence that patients in the intervention group were more likely to refill their prescriptions at 3 months than at 1 month.

In the intervention group, patients’ knowledge about ACS and their medications increased significantly from baseline to 1 day (after video viewing) and 3 months after ACS. Thus, the intervention appeared to effectively improve patients’ cardiovascular knowledge over the first 3 months after ACS presentation.

Overall, our results are encouraging but must be interpreted with caution. The sample size initially calculated could not be reached, and our study is therefore underpowered. However, our results are still interesting and show a great tendency for improvement of drug adherence and knowledge.

Comparison With Previous Work

App use has been demonstrated to improve drug adherence. For example, use of the “My Interventional Drug-Eluting Stent Education” app, developed in the United States with a similar goal as the “Mon Coeur, Mon BASIC” video, increased patients’ antiplatelet drug adherence and knowledge about PCI in a pilot study [27]. Use of another patient support app increased ticagrelor adherence and tended to improve cardiovascular-related lifestyle changes [28]. mHealth technologies seem to have an impact in drug adherence and should be part of future developments in the cardiology field.

Place of mHealth Technology for ACS Patients

Most currently available mHealth technologies were developed to support cardiovascular health; products include eHealth diaries [23], apps supporting cardiac rehabilitation [24,25], and complete e-learning platforms [26]. We are convinced that these strategies will make up a large part of future patient care, but their use will likely remain limited for a large proportion of current cardiac patients. The median age of our patients at admission was 59 years, and 20% (12/60) of patients reported low levels of ICT use. Moreover, we had to exclude patients with no ICT device or internet connection at home. For these reasons, we sought to develop a tool that is easy to use, even for patients who are not familiar with ICT use. We developed pocket-sized medication cards instead of digital cards for the same reasons, and patients appreciated these cards even more than the video. mHealth cannot replace the face-to-face approach, but it can be used as a complementary tool to increase patients' self-efficacy. In this study, the pharmacist talked with patients and answered their questions after they had watched the video, which may have affected their satisfaction. We will continuously update the content and enhance the interactivity of our web-based video, and we plan to develop a complete app with an easily updated medication plan, a frequently asked questions module, and quizzes to help patients develop even more knowledge.

Limitations

This study has some limitations. The ARMS is not the gold standard for the assessment of drug adherence; such assessment should include the direct measurement of drug consumption. Unfortunately, patients' drug consumption cannot be determined accurately using data from the Swiss health care system, as drugs are delivered by packs containing a prespecified number of tablets (eg, ticagrelor is available in blister packs of 56 tablets, whereas aspirin is delivered by packs of 98 tablets) and are not delivered by drug unit (for example, 30 tablets for 30 days of treatment). In addition, we did not adjust for some baseline characteristics (eg, age, sex, and educational level), which may have resulted in bias. Nevertheless, all baseline characteristics were similar in the two study groups, which reduced the risk of statistical errors; we thus believe that the intervention had an impact on drug adherence. Another limitation is the small sample, which together with the high degree of variance resulted in the loss of statistical power and the inability to perform subgroup analyses. We had difficulties in recruiting patients because many of them were transferred within 24 hours to a peripheral hospital, and we lost several of them to follow-up. We unfortunately could not continue the recruitment further, and we had to analyze the results with the small sample. The exclusion of patients without home internet access also may have biased the results. Finally, our findings may have been affected by attrition bias, as we lost about 20% (13/60) of patients between baseline and study completion. Despite these limitations, our results are very encouraging and suggest that the provision of an interactive informational video to patients has positive effects. These results should be confirmed in larger clinical studies that include subgroup analyses to identify the populations that would benefit most from such interventions.

Conclusions

Despite our study being underpowered, we were still able to show that the "Mon Coeur, Mon BASIC" web-hosted interactive video improved drug adherence and enhanced patients' cardiovascular knowledge. The video will be available at no charge from a webpage constructed by our hospital's cardiology department staff and will be offered to all patients hospitalized for ACS at our institution in the future. Our findings reflect patients' need for and appreciation of medical information; the most appropriate means of providing such information needs to be determined. Such tools should be developed for patients with a wide range of chronic diseases, and their content should be continuously improved. Over time, increasing numbers of patients will be able to use smartphone apps, which provides the opportunity to develop this type of support to improve the management of patients with chronic diseases.

Acknowledgments

This work was supported by an "e-learning commission" from the University Hospital of Lausanne. The funder had no role in the design of the study; collection, analysis, or interpretation of data; or writing of the manuscript. We sincerely thank all individuals who participated in the development of the "Mon Coeur, Mon BASIC" video, including Alienor Held, Pierre-Antoine Dubois, Ana Loureiro Soares, Anaëlle Morf, and Anne-Sylvie Diezi; we thank Marc Sohrmann for his advice on video content and Laurence Iseli for video narration.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Medication card distributed to the intervention group, in addition to the video viewing.

[DOCX File , 60 KB - jmir_v23i8e21938_app1.docx]

Multimedia Appendix 2

ARMS score questionnaire developed by Kripalani et al, with the presentation of the 12 questions and his scoring system. ARMS: Adherence to Refills and Medication Scale.

[DOCX File , 17 KB - jmir_v23i8e21938_app2.docx]

Multimedia Appendix 3

Questionnaire used for the assessment of the patient's knowledge. Correct answers were evaluated with 1 point and wrong answers with 0 points. Maximal points were 9 points.

[DOCX File , 18 KB - jmir_v23i8e21938_app3.docx]

Multimedia Appendix 4
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1066 KB - jmir_v23i8e21938_app4.pdf]

References

1. Nichols M, Townsend N, Scarborough P, Rayner M. Cardiovascular disease in Europe 2014: epidemiological update. Eur Heart J 2014 Nov 07;35(42):2929-2933. [doi: 10.1093/eurheartj/ehu378] [Medline: 25381246]
2. Mukherjee D, Fang J, Chetcuti S, Moscucci M, Kline-Rogers E, Eagle KA. Impact of combination evidence-based medical therapy on mortality in patients with acute coronary syndromes. Circulation 2004 Feb 17;109(6):745-749. [doi: 10.1161/01.CIR.0000112577.69066.CB] [Medline: 14970110]
3. Ibanez B, James S, Agewall S, Antunes MJ, Bucciarelli-Ducci C, Bueno H, ESC Scientific Document Group. 2017 ESC Guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation: The Task Force for the management of acute myocardial infarction in patients presenting with ST-segment elevation of the European Society of Cardiology (ESC). Eur Heart J 2018 Jan 07;39(2):119-177. [doi: 10.1093/eurheartj/ehx393] [Medline: 28886621]
4. Roffi M, Patrono C, Collet J, Mueller C, Valgimigli M, Andreotti F, ESC Scientific Document Group. 2015 ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation: Task Force for the Management of Acute Coronary Syndromes in Patients Presenting without Persistent ST-Segment Elevation of the European Society of Cardiology (ESC). Eur Heart J 2016 Jan 14;37(3):267-315. [doi: 10.1093/eurheartj/ehv320] [Medline: 26320110]
5. Bruggmann C, Iglesias JF, Gex-Fabry M, Fesselet R, Vogt P, Sadeghipour F, et al. Long-Term Quality of Prescription for ST-Segment Elevation Myocardial Infarction (STEMI) Patients: A Real World 1-Year Follow-Up Study. Am J Cardiovasc Drugs 2020 Feb;20(1):105-115 [FREE Full text] [doi: 10.1007/s40256-019-00361-5] [Medline: 31300969]
6. Melloni C, Alexander KP, Ou F, LaPointe NMA, Roe MT, Newby LK, et al. Predictors of early discontinuation of evidence-based medicine after acute coronary syndrome. Am J Cardiol 2009 Jul 15;104(2):175-181. [doi: 10.1016/j.amjcard.2009.03.013] [Medline: 19576342]
7. Kassab Y, Hassan Y, Abd Aziz N, Ismail O, AbdulRazzaq H. Patients' adherence to secondary prevention pharmacotherapy after acute coronary syndromes. Int J Clin Pharm 2013 Apr 3;35(2):275-280. [doi: 10.1007/s11096-012-9735-y] [Medline: 23283596]
8. Kolandaivelu K, Leiden BB, O'Gara PT, Bhatt DL. Non-adherence to cardiovascular medications. Eur Heart J 2014 Dec 07;35(46):3267-3276. [doi: 10.1093/eurheartj/ehu364] [Medline: 25265973]
9. Newby LK, LaPointe NMA, Chen AY, Kramer JM, Hammill BG, DeLong ER, et al. Long-term adherence to evidence-based secondary prevention therapies in coronary artery disease. Circulation 2006 Jan 17;113(2):203-212. [doi: 10.1161/CIRCULATIONAHA.105.505636] [Medline: 16401776]
10. Rasmussen JN, Chong A, Alter DA. Relationship between adherence to evidence-based pharmacotherapy and long-term mortality after acute myocardial infarction. JAMA 2007 Jan 10;297(2):177-186. [doi: 10.1001/jama.297.2.177] [Medline: 17213401]
11. Choudhry NK, Glynn RJ, Avorn J, Lee JL, Brennan TA, Reisman L, et al. Untangling the relationship between medication adherence and post-myocardial infarction outcomes: medication adherence and clinical outcomes. Am Heart J 2014 Jan;167(1):51-58.e5. [doi: 10.1016/j.ahj.2013.09.014] [Medline: 24332142]
12. Mehran R, Baber U, Steg PG, Ariti C, Weisz G, Witzenbichler B, et al. Cessation of dual antiplatelet treatment and cardiac events after percutaneous coronary intervention (PARIS): 2 year results from a prospective observational study. Lancet 2013 Nov 23;382(9906):1714-1722. [doi: 10.1016/S0140-6736(13)61720-1] [Medline: 24004642]
13. Osterberg L, Blaschke T. Adherence to medication. N Engl J Med 2005 Aug 04;353(5):487-497. [doi: 10.1056/NEJMra050100] [Medline: 16079372]
14. Clark PA, Drain M, Gesell SB, Mylod DM, Kaldenberg DO, Hamilton J. Patient perceptions of quality in discharge instruction. Patient Educ Couns 2005 Oct;59(1):56-68. [doi: 10.1016/j.pec.2004.09.010] [Medline: 16198219]
15. Booth K, Beaver K, Kitchener H, O'Neill J, Farrell C. Women's experiences of information, psychological distress and worry after treatment for gynaecological cancer. Patient Educ Couns 2005 Feb;56(2):225-232. [doi: 10.1016/j.pec.2004.02.016] [Medline: 15653253]
16. Prinjha S, Chapple A, Herxheimer A, McPherson A. Many people with epilepsy want to know more: a qualitative study. Fam Pract 2005 Aug;22(4):435-441. [doi: 10.1093/fampra/cmi024] [Medline: 15814581]
17. Murray MD, Young J, Hoke S, Tu W, Weiner M, Morrow D, et al. Pharmacist intervention to improve medication adherence in heart failure: a randomized trial. Ann Intern Med 2007 May 15;146(10):714-725. [doi: 10.7326/0003-4819-146-10-200705150-00005] [Medline: 17502632]

18. Smith DH, Kramer JM, Perrin N, Platt R, Roblin DW, Lane K, et al. A randomized trial of direct-to-patient communication to enhance adherence to beta-blocker therapy following myocardial infarction. Arch Intern Med 2008 Mar 10;168(5):477-83; discussion 483; quiz 447. [doi: 10.1001/archinternmed.2007.132] [Medline: 18332291]
19. Schwalm J, Ivers NM, Natarajan MK, Taljaard M, Rao-Melacini P, Witteman HO, et al. Cluster randomized controlled trial of Delayed Educational Reminders for Long-term Medication Adherence in ST-Elevation Myocardial Infarction (DERLA-STEMI). Am Heart J 2015 Nov;170(5):903-913 [FREE Full text] [doi: 10.1016/j.ahj.2015.08.014] [Medline: 26542498]
20. Budiman T, Snodgrass K, Komatsu Chang A. Evaluation of Pharmacist Medication Education and Post-discharge Follow-up in Reducing Readmissions in Patients With ST-Segment Elevation Myocardial Infarction (STEMI). Ann Pharmacother 2016 Feb 17;50(2):118-124. [doi: 10.1177/1060028015620425] [Medline: 26681443]
21. Calvert SB, Kramer JM, Anstrom KJ, Kaltenbach LA, Stafford JA, Allen LaPointe NM. Patient-focused intervention to improve long-term adherence to evidence-based medications: a randomized trial. Am Heart J 2012 Apr;163(4):657-65.e1. [doi: 10.1016/j.ahj.2012.01.019] [Medline: 22520532]
22. Agvall B, Alehagen U, Dahlström U. The benefits of using a heart failure management programme in Swedish primary healthcare. Eur J Heart Fail 2013 Feb 27;15(2):228-236 [FREE Full text] [doi: 10.1093/eurjhf/hfs159] [Medline: 23109650]
23. Wolf A, Fors A, Ulin K, Thorn J, Swedberg K, Ekman I. An eHealth Diary and Symptom-Tracking Tool Combined With Person-Centered Care for Improving Self-Efficacy After a Diagnosis of Acute Coronary Syndrome: A Substudy of a Randomized Controlled Trial. J Med Internet Res 2016 Feb 23;18(2):e40 [FREE Full text] [doi: 10.2196/jmir.4890] [Medline: 26907584]
24. Beatty AL, Magnusson SL, Fortney JC, Sayre GG, Whooley MA. VA FitHeart, a Mobile App for Cardiac Rehabilitation: Usability Study. JMIR Hum Factors 2018 Jan 15;5(1):e3 [FREE Full text] [doi: 10.2196/humanfactors.8017] [Medline: 29335235]
25. Frederix I, Hansen D, Coninx K, Vandervoort P, Vandijck D, Hens N, et al. Medium-Term Effectiveness of a Comprehensive Internet-Based and Patient-Specific Telerehabilitation Program With Text Messaging Support for Cardiac Patients: Randomized Controlled Trial. J Med Internet Res 2015 Jul 23;17(7):e185 [FREE Full text] [doi: 10.2196/jmir.4799] [Medline: 26206311]
26. Fontaine G, Cossette S, Heppell S, Boyer L, Mailhot T, Simard M, et al. Evaluation of a Web-Based E-Learning Platform for Brief Motivational Interviewing by Nurses in Cardiovascular Care: A Pilot Study. J Med Internet Res 2016 Aug 18;18(8):e224 [FREE Full text] [doi: 10.2196/jmir.6298] [Medline: 27539960]
27. Shah V, Dileep A, Dickens C, Groo V, Welland B, Field J, et al. Patient-Centered Tablet Application for Improving Medication Adherence after a Drug-Eluting Stent. Front Public Health 2016 Dec 12;4:272 [FREE Full text] [doi: 10.3389/fpubh.2016.00272] [Medline: 28018897]
28. Johnston N, Bodegard J, Jerström S, Åkesson J, Brorsson H, Alfredsson J, et al. Effects of interactive patient smartphone support app on drug adherence and lifestyle changes in myocardial infarction patients: A randomized study. Am Heart J 2016 Aug;178:85-94 [FREE Full text] [doi: 10.1016/j.ahj.2016.05.005] [Medline: 27502855]
29. ELIPS pour combattre l'infarctus et l'athérosclérose. Hôpitaux Universitaires de Genève. 2018. URL: https://www.hug-ge.ch/elips [accessed 2018-11-01]
30. CHUV service de cardiologie. Mon coeur, mon BASIC, l'essentiel à savoir après un infarctus. 2021 Jan 04. URL: https://www.chuv.ch/video-basic [accessed 2021-04-02]
31. Geneva University Hospitals. Pour combattre l'infarctus du myocarde et l'athérosclérose. 2010 Dec 23. URL: https://www.youtube.com/watch?v=49UUp_x30AY [accessed 2021-04-02]
32. Kripalani S, Risser J, Gatti ME, Jacobson TA. Development and evaluation of the Adherence to Refills and Medications Scale (ARMS) among low-literacy patients with chronic disease. Value Health 2009 Jan;12(1):118-123 [FREE Full text] [doi: 10.1111/j.1524-4733.2008.00400.x] [Medline: 19911444]
33. Sousa V, Rojjanasrirat W. Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline. J Eval Clin Pract 2011 Apr;17(2):268-274. [doi: 10.1111/j.1365-2753.2010.01434.x] [Medline: 20874835]
34. Ousseine YM, Rouquette A, Bouhnik A, Rigal L, Ringa V, Smith, et al. Validation of the French version of the Functional, Communicative and Critical Health Literacy scale (FCCHL). J Patient Rep Outcomes 2017 Feb 7;2(1):3 [FREE Full text] [doi: 10.1186/s41687-018-0027-8] [Medline: 29757319]
35. 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. [doi: 10.1016/j.jbi.2008.08.010]
36. 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 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]

Abbreviations

ACS: acute coronary syndrome
ARMS: Adherence to Refills and Medication Scale
BASIC: Secondary Prevention of ACS With Beta-Blockers, Antiaggregants, Statins, Angiotensin-Converting Enzyme Inhibitors, and Risk Factor Control
CR: cardio-rehabilitation
FCCHL: Functional, Communicative and Critical Health Literacy
ICT: information and communications technology
mHealth: mobile health
MI: myocardial infarction
NSTEMI: non–ST-segment elevation myocardial infarction
PCI: percutaneous coronary intervention
REDCap: Research Electronic Data Capture
STEMI: ST-segment elevation myocardial infarction

Edited by G Eysenbach; submitted 30.06.20; peer-reviewed by J Creutzfeldt, M Lozano-Lozano; comments to author 20.10.20; revised version received 11.03.21; accepted 03.04.21; published 30.08.21.

Please cite as:
Bruggmann C, Adjedj J, Sardy S, Muller O, Voirol P, Sadeghipour F
Effects of the Interactive Web-Based Video "Mon Coeur, Mon BASIC" on Drug Adherence of Patients With Myocardial Infarction: Randomized Controlled Trial
J Med Internet Res 2021;23(8):e21938
URL: https://www.jmir.org/2021/8/e21938
doi:10.2196/21938
PMID:34459744

©Christel Bruggmann, Julien Adjedj, Sylvain Sardy, Olivier Muller, Pierre Voirol, Farshid Sadeghipour. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 30.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Review

Ecological Momentary Assessment and mHealth Interventions Among Men Who Have Sex With Men: Scoping Review

Viktor Clark[1], BA, MS; Sunny Jung Kim[1], BA, MA, MS, PhD

Department of Health Behavior and Policy, School of Medicine, Virginia Commonwealth University, Richmond, VA, United States

Corresponding Author:
Viktor Clark, BA, MS
Department of Health Behavior and Policy
School of Medicine
Virginia Commonwealth University
830 East Main Street
Richmond, VA, 23219
United States
Phone: 1 8046283443
Email: nowackv@mymail.vcu.edu

Abstract

Background: Ecological momentary assessment (EMA) is a research design that allows for the measurement of nearly instantaneous experiences within the participant’s natural environment. Using EMA can help improve recall bias, ecological validity, and patient engagement while enhancing personalization and the ubiquity of interventions. People that can benefit from the use of EMA are men who have sex with men (MSM). Previous EMA studies have been successful in capturing patterns of depression, anxiety, substance use, and risky sexual behavior. These findings are directly relevant to MSM, who have high rates of each of these psychological and behavioral outcomes. Although there is a driving force behind the growing literature surrounding EMAs among MSM, no synthesizing reviews yet exist.

Objective: The aims of this study were to (1) synthesize the literature across fields on how EMA methods have been used among MSM, (2) better understand the feasibility and acceptability of EMA interventions among MSM, and (3) inform designs for future research studies on best evidence-based practices for EMA interventions.

Methods: Based on 4 library databases, we conducted a scoping review of EMAs used within interventions among MSM. The eligibility criteria included peer-reviewed studies conducted in the United States and the use of EMA methodology in an intervention for MSM. Modeling after the Centers for Disease Control and Prevention’s Compendium of Evidence-Based Interventions as the framework, we applied a typology that used 8 distinct review criteria, for example, sample size, design of the intervention, random assignment, design of the follow-up investigation, rate of retention, and rate of engagement.

Results: Our results (k=15, N=952) indicated a range of sample sizes; the smallest sample size was 12, while the largest sample size was 120. Of the 15 studies, 7 (47%) focused on outcomes related to substance use or outcomes related to psychological experiences. Of the 15 studies, 5 (33%) implemented an EMA intervention across 30 days. Of the 15 studies, 2 studies (13%) used random assignment, and 2 studies (13%) had quasi-experimental designs. Of the 15 studies, 10 studies (67%) reported acceptable retention rates greater than 70%. The outcomes that had event-contingent prompts (ie, prompts after engaging in substance use) were not as effective in engaging participants, with overall engagement rates as low as 37%.

Conclusions: Our systematic scoping review indicates strong evidence that the EMA methodology is both feasible and acceptable at high rates among MSM, especially, when examining psychological and behavioral outcomes such as negative or positive affect, risky sexual behavior, or substance use. Further research on optimal designs of EMA interventions for MSM is warranted.

(J Med Internet Res 2021;23(8):e27751) doi:10.2196/27751

KEYWORDS

mHealth; men who have sex with men; mobile health; interventions; mental health; sexual health; ecological momentary assessment; behavior

Introduction

Ecological Momentary Assessment

Developed originally from the field of personality and social psychology, ecological momentary assessment (EMA) is a research design with methodological components that allow researchers to measure experiences as close to that moment as possible and within the participant's natural environment [1]. Methodological strategies in EMAs have included prompting participants at various time intervals using self-report surveys or triggers specific to locations, events, or both [2]. Researchers have indicated that recall-based self-reports can be inaccurate or unreliable measurements of participants' actual lived experiences [2]. The goal of EMA is to minimize retrospective recall issues and enhance researchers' ability to measure lived experiences of people in the moment [1].

The ability to measure and potentially intervene in lived experiences in the moment is especially important to impact dysfunctional thoughts, capture psychological distress, or even intervene in harmful behavior [2]. In addition, one of the key benefits and goals of EMA is to provide high levels of ecological validity, or validity that comes from collecting data and implementing an experiment in a participant's natural setting in real-world contexts. High ecological validity can enhance the ability for research findings to be applied to real-world scenarios, increasing the likelihood of generalizability [3]. Research has also found EMA methods to outperform traditional paper-pencil measurements in the ability to determine needs of clinical interventions more precisely. A primary reason that EMA measurements outperform traditional paper-pencil measurements is that repeated measurements minimize the effect of participants' current state on results [4]. Finally, technology-based interventions incorporating EMA methods have shown promise in terms of feasibility and acceptability of enhancing intervention outcomes [5].

Men Who Have Sex With Men

Men who have sex with men (MSM) have been found to show high rates of both psychological distress and engagement in various risky behaviors [6]. Specifically, studies have found MSM to endure higher levels of depression, anxiety, substance use, and risk of contracting HIV [7,8]. EMAs have been used among MSM in daily diaries since 2007 [7] and have evolved tremendously into the realm of internet use [9], smartphone technology [10], and interventions [11]. Use of EMAs among MSM is a growing area of research. EMAs have been shown to be highly effective in reaching people who have a history of substance use or other risky behavior, due to the minimization of stigma and enhancement of self-control over privacy, confidentiality, and anonymity [10,12,13].

Scoping Review

The primary purpose of a scoping review is to synthesize current literature surrounding a topic area. Thus, the synthesis produced from a scoping review acts as a summary of available literature, a means to identify key concepts, and a precursor to a systematic review [14]. To the best of the authors' knowledge, neither a scoping review nor a systematic review has yet been produced on the topic of EMA use among MSM, due to the limited literature surrounding the topic. As a result, the authors intended to conduct a scoping review by applying the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses-Extension for Scoping Reviews) [15], as a contribution toward a future systematic review based on an increase in literature. Authors also acknowledge the limitations of a scoping review, mainly, an inability to make quality assurances, a lack of strong validity, and an inability to hypothesize based on the review.

Scoping Review for EMAs Among MSM

Despite the benefits and clinical implications of using EMA methods for at-risk populations, there have been no reviews compiling the literature of how EMAs have been used among MSM. One growing method for synthesizing theoretical and empirical evidence in the literature is the scoping review [16]. Scoping reviews are considered as means to describe key findings across literature, identify gaps in research, and inform the design of future research studies [16]. Two major benefits of conducting a scoping review are the ability to examine the breadth of the topic of EMA methodologies that are applied to MSM, specifically, within interventions, and the ability to identify knowledge gaps and future directions for the expansion of this area of research. The aims of this study were to: (1) synthesize the literature across fields on how EMA methods have been used among MSM, (2) better understand the feasibility and acceptability of EMA interventions among MSM, and (3) to inform designs for future research studies on best evidence-based practices for EMA interventions.

Methods

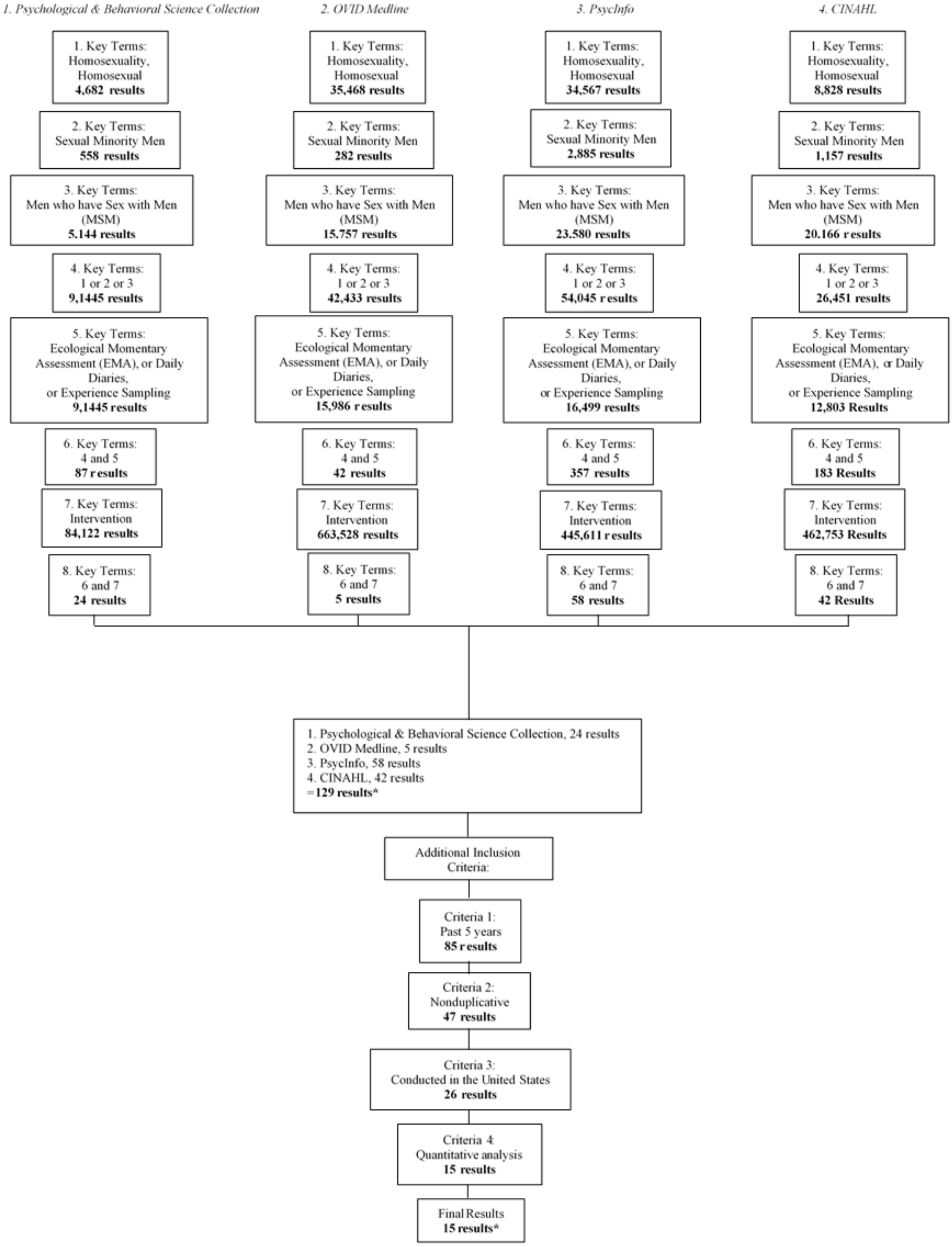
Eligibility Criteria

We identified several eligibility criteria that needed to be met for a publication to be included in this scoping review. The eligibility criteria included MSM samples; EMA interventions or assessments or determinations of EMA's efficacy, acceptability, and feasibility; a publication date within the past 5 years; peer-reviewed studies conducted in the United States; and quantitative data analyses.

Information Sources

We chose 4 prominent databases to retain studies from: Ovid Medline, which focuses on biomedical scholarly literature; Psychological and Behavioral Science Collection, which focuses on mental processes and emotional and behavioral experiences; PsycInfo, which focuses on behavioral and social science research; and Cumulative Index to Nursing and Allied Health Literature, which provides access to health research, specifically, in nursing and other allied health. EMAs have predominately been applied to behavioral or psychological health [17,18] and physical health [19,20]. Therefore, these databases were determined as most relevant and applicable to the specific topic of this scoping review (Figure 1).

Figure 1. Flow diagram of study selection process. CINAHL: Cumulative Index to Nursing and Allied Health Literature; *Peer-reviewed articles.


1. Psychological & Behavioral Science Collection
2. OVID Medline
3. PsycInfo
4. CINAHL
1. Key Terms: Homosexuality, Homosexual 4,682 results
1. Key Terms: Homosexuality, Homosexual 35,468 results
1. Key Terms: Homosexuality, Homosexual 34,567 results
1. Key Terms: Homosexuality, Homosexual 8,828 results
2. Key Terms: Sexual Minority Men 558 results
2. Key Terms: Sexual Minority Men 282 results
2. Key Terms: Sexual Minority Men 2,885 results
2. Key Terms: Sexual Minority Men 1,157 results
3. Key Terms: Men who have Sex with Men (MSM) 5.144 results
3. Key Terms: Men who have Sex with Men (MSM) 15.757 results
3. Key Terms: Men who have Sex with Men (MSM) 23.580 results
3. Key Terms: Men who have Sex with Men (MSM) 20.166 results
4. Key Terms: 1 or 2 or 3 9,1445 results
4. Key Terms: 1 or 2 or 3 42,433 results
4. Key Terms: 1 or 2 or 3 54,045 results
4. Key Terms: 1 or 2 or 3 26,451 results
5. Key Terms: Ecological Momentary Assessment (EMA), or Daily Diaries, or Experience Sampling 9,1445 results
5. Key Terms: Ecological Momentary Assessment (EMA), or Daily Diaries, or Experience Sampling 15,986 results
5. Key Terms: Ecological Momentary Assessment (EMA), or Daily Diaries, or Experience Sampling 16,499 results
5. Key Terms: Ecological Momentary Assessment (EMA), or Daily Diaries, or Experience Sampling 12,803 Results
6. Key Terms: 4 and 5 87 results
6. Key Terms: 4 and 5 42 results
6. Key Terms: 4 and 5 357 results
6. Key Terms: 4 and 5 183 Results
7. Key Terms: Intervention 84,122 results
7. Key Terms: Intervention 663,528 results
7. Key Terms: Intervention 445,611 results
7. Key Terms: Intervention 462,753 Results
8. Key Terms: 6 and 7 24 results
8. Key Terms: 6 and 7 5 results
8. Key Terms: 6 and 7 58 results
8. Key Terms: 6 and 7 42 Results
1. Psychological & Behavioral Science Collection, 24 results
2. OVID Medline, 5 results
3. PsycInfo, 58 results
4. CINAHL, 42 results
=129 results*
Additional Inclusion Criteria:
Criteria 1: Past 5 years 85 results
Criteria 2: Nonduplicative 47 results
Criteria 3: Conducted in the United States 26 results
Criteria 4: Quantitative analysis 15 results
Final Results 15 results*

Keywords and Search Process

To ensure that we conducted an inclusive review of the literature across databases, we used multiple search terms for our target populations. Step 1 involved searching the terms "homosexuality" and "homosexual." In step 2, we used the search term "sexual minority men." Finally, in step 3, we used the search terms "men who have sex with men" and "MSM." In step 4, we combined all these terms to achieve the most comprehensive review of literature pertaining to our target population of MSM. In step 5, we added the keywords "ecological momentary assessment" or "EMA," "daily diaries," and "experience sampling" to identify all EMA-related literature from these databases. In step 6, we combined our MSM search terms with the EMA terms to narrow down to only the most directly relevant articles. Upon reaching step 7, we applied the term "intervention" to further narrow the articles included in this review. Step 8 was comprised of combining all search terms from our MSM terms, EMA terms, and intervention terms, which resulted in a total of 129 articles.

Inclusion Criteria, Exclusion Criteria, and the Iterative Process

Once we had the initial studies from our database searches, we combined all the study titles and previewed the articles. Articles were further narrowed based on 2 additional inclusion criteria and 2 additional exclusion criteria; studies were excluded if they were not published in the past 5 years and if they were duplicative across databases. Studies were included if they were conducted in the United States and included quantitative data and analyses (Figure 1).

As suggested by the clarity of guidelines in scoping reviews [21], our search and review were conducted in an iterative manner over time. We conducted our first search in January 2020, our second search in May 2020, our third search in September 2020, and the last search in March 2021, as presented, to examine quarterly changes in the literature.

Analysis

Modeling after the Centers for Disease Control and Prevention (CDC) Compendium of Evidence-Based Interventions (EBIs) [22], we primarily used 8 distinct criteria to review the final set of eligible publications [9,11,13,23-34]. The EBI criteria and best practices in the compendium, developed and defined by the CDC's Prevention Research Synthesis project, posit a series of systematic review components for interventions. EBI criteria have been shown to generate significant effects and strong evidence of efficacy in HIV-related outcomes [22]. Based on the compendium of EBIs, our review criteria included (1) citation, sample size, and duration of study; (2) location; (3) random assignment (yes or no); (4) key aspects of the intervention; (5) follow-up (yes or no); (6) occurrence of follow-up after intervention; (7) rate of retention; and (8) rate of engagement (Multimedia Appendix 1). We also conducted a secondary analysis comprised of additional review criteria: recruitment strategy, description of intervention and comparison of arms, specific measurement tools, outcomes of interest, and outcome results (Multimedia Appendix 2).

Results

Risk of Quality in Individual Studies

Using the CDC compendium of best evidence-based risk reduction for individual-level interventions, we evaluated the strength of each study. None of the studies reviewed met full criteria for best standards of risk reduction in individual-level interventions. However, we continued evaluating based off CDC compendium criteria to determine study designs, intervention elements, and highest standards currently achieved.

Study Selection

A total of 129 articles were identified in the preliminary search: 24 from the Psychology and Behavioral Science Collection, 5 from OVID Medline, 58 from PsycInfo, and 42 from Cumulative Index to Nursing and Allied Health Literature. Among these 129 articles, we excluded 114 studies for not meeting additional screening criteria. These 4 criteria that the studies had to meet were: (1) the study was published in the past 5 years, eliminating 44 studies; (2) not a duplicative study, eliminating 38 studies; (3) a US-based study, eliminating 12 studies; and (4) the study used a quantitative analysis, eliminating 9 studies. The qualifying criteria led to a final set of 15 studies in this review [9,11,13,23-34].

Sample Sizes

The average sample size across the 15 studies reviewed was 63.5 (SD 31.85). The CDC recommends that each sample is greater than 50 participants per study arm. Among the 15 studies reviewed, 10 (67%) had sample sizes >50, and 4 (27%) had sample sizes ≥100 [9,31,33,34]. The largest sample size was 120, in a study that had a single intervention arm [31]. Of 15 studies, 4 (27%) had multiple intervention arms [11,26,27,30].

Demographics

The only 2 demographic measurements reported by all 15 articles were race/ethnicity and age [9,11,13,23-34]. Within the total sample of 952 participants that was developed from a composite of all article samples, the majority of participants (476/952, 50%) were white, and this sample had a mean age of 38.75 years (SD 8.5). Of 15 studies, across 9 studies (60%), income was also reported, with the majority of the composite sample making <$40,000 annually [9,11,13,25,26,31-34]. The final demographic measurement that was majorly reported, across 12 of the 15 studies (80%), was education; the majority (420/952, 44.1%) of this composite sample had at least some college education [9,11,13,24-26,28,29,31-34]. A detailed synopsis of demographics can be found in Multimedia Appendix 3.

Key Aspects of the Interventions

From this review, we determined that multiple studies had similar key aspects across the interventions implemented. First, 7 of the 15 studies (47%) focused on multiple types of substance use, including nicotine use, alcohol use, or other substance use (eg, cocaine, methamphetamine, and cannabis). Second, affect and stigma were discussed as primary outcomes for 5 of 15 studies (33%) [11,28-30,34] and secondary outcomes for 3 additional studies of the 15 studies (20%) [27,32,33]. Among 15 studies reviewed, 6 (40%) of them focused on examining feasibility or acceptability of EMA methodologies within the intervention [11,13,23,25,31]. Finally, 60% (9/15) of the studies focused on sexual behavior among men who have sex with men [9,11,23,25,28,30,32-34].

Random Assignment

Random assignment occurred in 2 of the 15 studies (13%) included in our review [11,26]. The CDC recommends random assignment as a gold standard, to rule out biases in a systematic way across multiple intervention arms. Since the majority of the studies (11/15, 73%) included in this review had only 1 arm, random assignment was not implemented [9,13,23-25,28,29,32-34]. In the remaining 2 studies that had nonrandomized designs with multiple arms, one study used a quasi-experimental design implementing clinical cutoffs for hypersexuality to determine group membership [30], while the other study assigned groups based on whether participants were recruited in-person or online [9]. Random assignment was considered as potentially unethical in many of these studies due to their focus on substance use and sexual behavior; thus, a quasi-experimental design was better suited.

Description of Intervention and Comparison Arms

Across the articles, there were a wide variety of intervention strategies used for implementation, for example, impacting the duration, mechanism for the intervention, and tasks for intervention participants. The average duration of the intervention was 31.5 days; the shortest intervention was a single-day cross-sectional survey [13], and the longest was a 90-day intervention [31]. Interventions were conducted primarily via mobile devices (14/15, 94%) [9,11,23-34]. Among the interventions conducted through mobile devices, 11 of 14 studies (79%) used texting prompts [13,28-30]. Tasks to be completed were predominately surveys (10/15, 67%) [11,13,24-31] administered through text message (6/15, 40%) [11,24-27,31]. The second-most used modality for data collection was daily diaries (5/15, 33%) [9,23,32-34]. The CDC recommends a clear intervention description, which we found across all the included studies [22].

Follow-up and Occurrence of Follow-up

Follow-up assessments were administered in 2 studies of the 15 (13%) [11,31], one of which included follow-ups at 3 different time points: 60 days, 90 days, and 120 days [31], while the other followed up with participants after 4 weeks [11]. According to the CDC compendium, there should be a follow-up, and it should, specifically, occur more than 30 days after completion of the intervention [22].

Rate of Retention

Rate of retention was measured in 33% (5/15) of the included studies [11,25-27,31]. Among those that measured retention, the average retention rate was 77.58%, the lowest retention rate was 29.2% [31], and the highest retention rate was 93% [26]. For a high-quality intervention, the CDC recommends a 70% study retention rate [22].

Rate of Engagement

Of the 15 studies, 13 (87%) reported rate of engagement [9,13,23-29,31-34]. Engagement rate was defined as an overall rate of completion for text or online surveys, text prompts, or daily diaries, depending on the study modality and was reported consistently across all studies. The average overall engagement rate was 76.93%, the lowest overall engagement rate was 37.3% [9], and the highest was 98.7% [13]. Engagement is a key component of retention, and the 70% retention rate is recommended as the benchmark for an acceptable engagement rate [22].

Location

The location of the catchment areas and study sites varied. The majority of studies (9/15, 60%) were conducted on the East Coast [9,13,23,25,28,29,32-34]. Within those conducted on the East Coast, the majority of these studies (5/9, 56%) were concentrated in the Northeast [9,23,32-34]. The second-most researched area was the West Coast (4/15, 27%) [11,26,27,31], with a focus on San Francisco (3/15, 20%) [26,27,31]. Of the 15 studies, 1 (7%) was conducted in the Northwest [24] and 1 (7%), in the upper Midwest [29].

Recruitment Strategies

Of the 15 studies, 12 (80%) used more than 2 recruitment strategies. Of those that used at least 2 recruitment strategies, 14/15 (93%) studies paired their strategies with social media (e.g., Instagram or Facebook) [9,11,23-34]. Of the 15 studies, 3 (20%) used cohorts from larger or alternative study sites [11,13,30], one of which used multiple recruitment strategies, including social media [30]. The most popular recruitment strategy that was paired with social media was the use of community-based organizations with in-person recruitment (6/15, 40%) [11,25-27,30,31].

Specific Measurement Tools

The majority of the studies (10/15, 67%) were conducted with EMA surveys [11,13,24-31]. Of these 10 studies, 2 (20%) used the same scale to measure affect, the Positive and Negative Affect Scale [9,30]. Of these 10 studies, 2 others (20%) administered the Difficulties with Emotion Regulation Scale to measure emotion dysregulation [28,29]. Many studies used questions such as how many partners a participant engaged in sex with over the past 24 hours (9/15, 60%) [9,11,23,25,30-34], how many standard alcoholic drinks a participant consumed in the past 24 hours (6/15, 40%) [9,24,26,32-34], and what types of drugs were used over the past 24 hours (11/15, 73%) [9,23,24,26,27,29-34].

Outcomes of Interest

There were 3 prominent outcomes of interest across the included studies: risky sexual behavior, substance use, and acceptability. Of the 15 studies, 7 (47%) measured substance use status, including use of nicotine, alcohol, and other nonprescription drugs [9,11,24,27,29,33,34]. Of the 15 studies, 6 (40%) measured sexual behavior, especially, risky, unprotected sexual behavior defined by condomless sex or sex with partners who were of unknown HIV status [11,23,30,32-34]. Of the 15 studies, 4 (27%) assessed the acceptability of EMAs implemented in an intervention by measuring response rates, completion rates, and study retention rates [9,25,26,31].

Outcome Results

Intervention studies using EMA methods have demonstrated success in longitudinally measuring substance use, compared with studies that relied on non-EMA measurements such as timeline follow-back surveys [9,23,32]. Additionally, EMA methods generated greater acceptability than other methods: daily diaries had high rates of response completion. The highest response rate was 97.3% [9,34], and the lowest was only 84% [23].

Discussion

Principal Findings

In our scoping review, we aimed to provide an overview of the growing literature on a relatively novel measurement: ecological momentary assessment (EMA). We found that among men who have sex with men (MSM), the majority of EMAs have been used to intervene on risk-taking behaviors such as alcohol and drug use or unprotected sex with multiple partners. Although risk-taking behaviors have often been stigmatized, the use of EMA through smartphone technology has been seen as a highly effective way to safely assess risk-taking behaviors [10,12,13]. Overall, EMA was seen as an acceptable and feasible method, with daily diaries as the most acceptable tool [9,23,32-34] to collect the experiences of MSM. A unique facet of MSM research was the successful use of recruitment strategies beyond technology-based recruitment, which included assistance in initiation and engagement from community-based organizations [11,25-27,30,31]. This scoping review was used as a synthesizing method with a wide array of review dimensions and criteria such as quantitative interventions among MSM. This allowed our results to provide a comprehensive set of typological frameworks that may be useful in designing and implementing an EMA-integrated intervention for behavioral change. Basing typographic dimensions off preexisting frameworks offered by the CDC [22], we incorporated the most salient components for intervention research. This allowed for a better assessment of the strength of existing EMA interventions among MSM. Also, we conducted 4 time points of literature searches (ie, January 2020, June 2020, September 2020, and March 2021), for inclusion of more studies, which increased the comprehensiveness of this scoping review.

Our study determined that there were limited psychometrically sound EMA measurements that were fully validated. Given the growing research attention on EMAs within the context of behavioral intervention, future studies may aim to develop and validate EMA measurements. EMA has been widely used as a just-in-time assessment and monitoring tool, but it also can be a great measurement resource to predict behavioral outcomes. We suggest future research should focus on developing predictive models and analytic methods, using intensive longitudinal data from EMAs to understand behavioral changes or outcomes over time. In terms of an analytic perspective, since EMAs lead to extensive longitudinal data, the risk of missingness and the handling of missing data will become more prevalent. Therefore, studies on appropriate analytic approaches to manage missing data from EMAs will be essential.

Although EMAs may reduce recall bias, due to the repetitiveness of measurement, they can also increase participant bias and burden [35,36]. Future research should take into consideration EMA designs that are engaging but protective of data anonymity and confidentiality, to prevent participant biases such as the social desirability effect and the halo or devil effect. To avoid priming of such participant biases, details and information in EMA-based interventions should be presented in a judgment-free manner. In order to reduce psychological reactions, future research should consider developing EMAs as self-motivated mechanisms, with use options such as event-contingent prompts, daily diaries, text prompts, or other mechanisms. Additionally, future studies should examine putative mechanism factors such as resilience and social support, to develop a comprehensive, integrative intervention program for MSM [37].

Limitations

There were limitations imposed by the scope and design of the study. First, the inclusion of the major library databases focused on studies most relevant to the population and methodological strategy of interest, but we excluded other minor library databases. Therefore, a future direction may include an inclusion of minor databases in the review for EMA interventions among MSM. Second, many of the studies included in this review were feasibility and acceptability tests as well as pilot studies with inconsistent assessments of outcomes, thus minimizing the effectiveness of a meta-analysis or systematic review. Although we presented results from a scoping review to provide an overview and the state of EMA in behavioral medicine, future research may conduct a systematic review or meta-analysis, as the prevalence of empirical evidence from randomized controlled trials using EMAs in this area is likely to increase [38].

Conclusions

Leveraging evidence-based intervention designs with validated ecological momentary assessments can advance our understanding of factors and processes in behavioral changes and health outcomes. These approaches can be further empowered through technology-based behavioral medicine and social medicine. In this scoping review paper, we provided a typology of EMA-based intervention research that was designed to promote health behavior and psychological well-being. Advancements in psychometric tests to validate EMAs will be critical. As the empirical evidence and theories in this field are emerging, we hope our review offers some guidance and synthesis of the literature to develop and evaluate technology-based EMA health interventions.

Authors' Contributions

VC conceived the study. VC and SJK were responsible for the study design and concept. SJK was responsible for accurate implementation of scoping review methodology. VC was responsible for the literature search and literature review. SJK supervised the study design and the review process. VC was responsible for constructing figures and multimedia appendices. VC and SJK collaborated on the discussion and study impact. VC and SJK drafted the first draft of the manuscript. All authors extensively reviewed and approved the manuscript before submission for peer review.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Primary scoping review analysis.
[DOCX File , 1380 KB - jmir_v23i8e27751_app1.docx]

Multimedia Appendix 2
Secondary scoping review analysis.
[DOCX File , 4852 KB - jmir_v23i8e27751_app2.docx]

Multimedia Appendix 3
Demographics.
[DOCX File , 21 KB - jmir_v23i8e27751_app3.docx]

References

1. Moskowitz DS, Young SN. Ecological momentary assessment: what it is and why it is a method of the future in clinical psychopharmacology. J Psychiatry Neurosci 2006 Jan;31(1):13-20 [FREE Full text] [Medline: 16496031]
2. Shiffman S, Stone AA, Hufford MR. Ecological Momentary Assessment. Annu. Rev. Clin. Psychol 2008 Apr;4(1):1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415]
3. Thong MSY, Chan RJ, van den Hurk C, Fessele K, Tan W, Poprawski D, et al. Going beyond (electronic) patient-reported outcomes: Harnessing the benefits of smart technology and ecological momentary assessment in cancer survivorship research. Support Care Cancer 2020 Aug 25;29(1):7-10. [doi: 10.1007/s00520-020-05648-x]
4. Moore RC, Depp CA, Wetherell JL, Lenze EJ. Ecological momentary assessment versus standard assessment instruments for measuring mindfulness, depressed mood, and anxiety among older adults. Journal of Psychiatric Research 2016 Apr;75:116-123. [doi: 10.1016/j.jpsychires.2016.01.011]
5. Kumar D, Tully LM, Iosif A, Zakskorn LN, Nye KE, Zia A, et al. A mobile health platform for clinical monitoring in early psychosis: Implementation in community-based outpatient early psychosis care. JMIR Ment Health 2018 Feb 27;5(1):e15. [doi: 10.2196/mental.8551]
6. Batchelder AW, Safren S, Mitchell AD, Ivardic I, O'Cleirigh C. Mental health in 2020 for men who have sex with men in the United States. Sex. Health 2017;14(1):59. [doi: 10.1071/sh16083]
7. HIV and gay and bisexual men. Centers for Disease Control Prevention. 2020. URL: https://www.cdc.gov/hiv/group/msm/index.html [accessed 2021-07-23]
8. Mustanski BS. Are sexual partners met online associated with HIV/STI risk behaviours? Retrospective and daily diary data in conflict. AIDS Care 2007 Jun 14;19(6):822-827. [doi: 10.1080/09540120701237244]
9. Carr DJ, Adia AC, Wray TB, Celio MA, Pérez AE, Monti PM. Using the internet to access key populations in ecological momentary assessment research: Comparing adherence, reactivity, and erratic responding across those enrolled remotely versus in-person. Psychological Assessment 2020 Aug;32(8):768-779. [doi: 10.1037/pas0000847]
10. Yang C, Linas B, Kirk G, Bollinger R, Chang L, Chander G, et al. Feasibility and acceptability of smartphone-based ecological momentary assessment of alcohol use among African American men who have sex with men in Baltimore. JMIR mHealth uHealth 2015 Jun 17;3(2):e67. [doi: 10.2196/mhealth.4344]
11. Reback CJ, Rünger D, Fletcher JB, Swendeman D. Ecological momentary assessments for self-monitoring and counseling to optimize methamphetamine treatment and sexual risk reduction outcomes among gay and bisexual men. Journal of Substance Abuse Treatment 2018 Sep;92:17-26. [doi: 10.1016/j.jsat.2018.06.005]
12. Singh NB, Björling EA. A review of EMA assessment period reporting for mood variables in substance use research: Expanding existing EMA guidelines. Addictive Behaviors 2019 Jul;94:133-146. [doi: 10.1016/j.addbeh.2019.01.033]
13. Duncan DT, Kapadia F, Kirchner TR, Goedel WC, Brady WJ, Halkitis PN. Acceptability of ecological momentary assessment among young men who have sex with men. Journal of LGBT Youth 2017 Sep 30;14(4):436-444. [doi: 10.1080/19361653.2017.1365038]
14. Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol 2018 Nov 19;18(1):143. [doi: 10.1186/s12874-018-0611-x]
15. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, Tunçalp, 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]
16. Lockwood C, dos Santos KB, Pap R. Practical guidance for knowledge synthesis: Scoping review methods. Asian Nursing Research 2019 Dec;13(5):287-294. [doi: 10.1016/j.anr.2019.11.002]
17. Gromatsky M, Sullivan SR, Spears AP, Mitchell E, Walsh S, Kimbrel NA, et al. Ecological momentary assessment (EMA) of mental health outcomes in veterans and servicemembers: A scoping review. Psychiatry Research 2020 Oct;292:113359. [doi: 10.1016/j.psychres.2020.113359]
18. Shiffman S. Ecological momentary assessment (EMA) in studies of substance use. Psychological Assessment 2009 Dec;21(4):486-497. [doi: 10.1037/a0017074]

19. Kampshoff CS, Verdonck-de Leeuw IM, Oijen MG, Sprangers MA, Buffart LM. Ecological momentary assessments among patients with cancer: A scoping review. Eur J Cancer Care 2019 May 14;28(3):1-11. [doi: 10.1111/ecc.13095]
20. Goldschmidt AB, Smith KE, Crosby RD, Boyd HK, Dougherty E, Engel SG, et al. Ecological momentary assessment of maladaptive eating in children and adolescents with overweight or obesity. Int J Eat Disord 2018 Apr 06;51(6):549-557. [doi: 10.1002/eat.22864]
21. Colquhoun HL, Levac D, O'Brien KK, Straus S, Tricco AC, Perrier L, et al. Scoping reviews: time for clarity in definition, methods, and reporting. Journal of Clinical Epidemiology 2014 Dec;67(12):1291-1294. [doi: 10.1016/j.jclinepi.2014.03.013]
22. HIV/AIDS prevention research synthesis project: Compendium evidence-based interventions and best practices for HIV prevention. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/hiv/research/interventionresearch/compendium/index.html [accessed 2021-07-23]
23. Wray TB, Kahler CW, Monti PM. Using Ecological Momentary Assessment (EMA) to study sex events among very high-risk men who have sex with men (MSM). AIDS Behav 2016 Jan 8;20(10):2231-2242. [doi: 10.1007/s10461-015-1272-y]
24. Livingston NA, Flentje A, Heck NC, Szalda-Petree A, Cochran BN. Ecological momentary assessment of daily discrimination experiences and nicotine, alcohol, and drug use among sexual and gender minority individuals. Journal of Consulting and Clinical Psychology 2017 Dec;85(12):1131-1143. [doi: 10.1037/ccp0000252]
25. Smiley SL, Elmasry H, Webb Hooper M, Niaura RS, Hamilton AB, Milburn NG. Feasibility of ecological momentary assessment of daily sexting and substance use among young adult African American gay and bisexual men: A pilot study. JMIR Res Protoc 2017 Feb 02;6(2):e9. [doi: 10.2196/resprot.6520]
26. Turner CM, Coffin P, Santos D, Huffaker S, Matheson T, Euren J, et al. Race/ethnicity, education, and age are associated with engagement in ecological momentary assessment text messaging among substance-using MSM in San Francisco. Journal of Substance Abuse Treatment 2017 Apr;75:43-48. [doi: 10.1016/j.jsat.2017.01.007]
27. Nguyen N, McQuoid J, Ramo D, Holmes LM, Ling PM, Thrul J. Real-time predictors of smoking among sexual minority and heterosexual young adults: An ecological momentary assessment study. Drug and Alcohol Dependence 2018 Nov;192:51-58. [doi: 10.1016/j.drugalcdep.2018.07.021]
28. Rendina HJ, Millar BM, Parsons JT. The critical role of internalized HIV-related stigma in the daily negative affective experiences of HIV-positive gay and bisexual men. Journal of Affective Disorders 2018 Feb;227:289-297. [doi: 10.1016/j.jad.2017.11.005]
29. Rendina HJ, Millar BM, Parsons JT. Situational HIV stigma and stimulant use: A day-level autoregressive cross-lagged path model among HIV-positive gay and bisexual men. Addictive Behaviors 2018 Aug;83:109-115. [doi: 10.1016/j.addbeh.2018.01.003]
30. Miner MH, Dickenson J, Coleman E. Effects of emotions on sexual behavior in men with and without hypersexuality. Sexual Addiction & Compulsivity 2019 Feb 07;26(1-2):24-41. [doi: 10.1080/10720162.2018.1564408]
31. Turner CM, Arayasirikul S, Trujillo D, Lê V, Wilson EC. Social inequity and structural barriers to completion of ecological momentary assessments for young men who have sex with men and trans women living with HIV in San Francisco. JMIR Mhealth Uhealth 2019 May 08;7(5):e13241. [doi: 10.2196/13241]
32. Wray TB, Luo X, Ke J, Pérez AE, Carr DJ, Monti PM. Using smartphone survey data and machine learning to identify situational and contextual risk factors for HIV risk behavior among men who have sex with men who are not on PrEP. Prev Sci 2019 May 9;20(6):904-913. [doi: 10.1007/s11121-019-01019-z]
33. Wray TB, Monti PM. Characteristics of sex events, partners, and motivations and their associations with HIV-risk behavior in a daily diary study of high-risk men who have sex with men (MSM). AIDS Behav 2020 Jun 24;24(6):1851-1864. [Medline: https://pubmed.ncbi.nlm.nih.gov/31832855/]
34. Wray TB, Monti PM, Kahler CW, Guigayoma JP. Using ecological momentary assessment (EMA) to explore mechanisms of alcohol‐involved HIV risk behavior among men who have sex with men (MSM). Addiction 2020 Apr 13;115(12):2293-2302. [doi: 10.1111/add.15053]
35. McCambridge J, de Bruin M, Witton J. The effects of demand characteristics on research participant behaviours in non-laboratory settings: A systematic review. PLoS ONE 2012 Jun 19;7(6):e39116. [doi: 10.1371/journal.pone.0039116]
36. Gove WR, Geerken MR. Response bias in surveys of mental health: An empirical investigation. American Journal of Sociology 1977 May;82(6):1289-1317. [doi: 10.1086/226466]
37. Perrin PB, Sutter ME, Trujillo MA, Henry RS, Pugh M. The minority strengths model: Development and initial path analytic validation in racially/ethnically diverse LGBTQ individuals. J. Clin. Psychol 2019 Aug 30;76(1):118-136. [doi: 10.1002/jclp.22850]
38. Doherty K, Balaskas A. The design of ecological momentary assessment technologies. Interacting with Computers 2020 May;32(3):257-278. [doi: 10.1093/iwcomp/iwaa019]

Abbreviations

CDC: Centers for Disease Control and Prevention
EBI: evidence-based intervention
EMA: ecological momentary assessment

MSM: men who have sex with men
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses - Extension for Scoping Reviews

Edited by R Kukafka; submitted 06.02.21; peer-reviewed by G Mansergh, C Jonassaint; comments to author 02.03.21; revised version received 07.05.21; accepted 24.05.21; published 03.08.21.

Please cite as:
Clark V, Kim SJ
Ecological Momentary Assessment and mHealth Interventions Among Men Who Have Sex With Men: Scoping Review
J Med Internet Res 2021;23(8):e27751
URL: https://www.jmir.org/2021/8/e27751
doi:10.2196/27751
PMID:34342585

©Viktor Clark, Sunny Jung Kim. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 03.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Gaps and Future Challenges of Italian Apps for Pregnancy and Postnatal Care: Systematic Search on App Stores

Laura Brunelli[1,2], MD; Chiara De Vita[3], MSc, PhD; Fabrizio Cenedese[3], MSc; Michela Cinello[3], MSc; Marta Paris[3], MSc; Francesca Samogizio[3], MSc; Anja Starec[3], MSc; Michele Bava[4], MSc, PhD; Margherita Dal Cin[5], MD; Sara Zanchiello[3], MSc; Tamara Stampalija[1,6], MD, PhD, Prof Dr

[1]Department of Medical, Surgical and Health Sciences, University of Trieste, Trieste, Italy
[2]Quality and Accreditation Unit, Friuli Centrale Healthcare University Trust, Udine, Italy
[3]Innovation and Complex Systems Department, Area Science Park, Trieste, Italy
[4]Clinical Engineering, ICT and Procurement, Institute for Maternal and Child Health, IRCCS Burlo Garofolo, Trieste, Italy
[5]Clinical Management Staff, Institute for Maternal and Child Health, IRCCS Burlo Garofolo, Trieste, Italy
[6]Unit of Fetal Medicine and Prenatal Diagnosis, Institute for Maternal and Child Health, IRCCS Burlo Garofolo, Trieste, Italy

Corresponding Author:
Laura Brunelli, MD
Department of Medical, Surgical and Health Sciences
University of Trieste
Via dell'Istria, 65
Trieste, 34137
Italy
Phone: 39 0432554768
Email: laura.brunelli@phd.units.it

Abstract

Background: Despite the availability of thousands of health apps worldwide, when considering those addressing children's first 1000 days of life, most apps fail to consider the continuity between the prenatal and postnatal stages, and their joint impact on maternal and child health. The reliability, quality, and effectiveness of these apps are largely unknown, and the provided content seems questionable in terms of completeness, updating, and trustworthiness.

Objective: This study evaluates available Italian pregnancy and postnatal care apps to highlight the main gaps to be overcome and the resulting future challenges to be met in this mobile health–related field.

Methods: A systematic search was conducted on the Apple App Store and Google Play Store, and basic information was collected for all identified apps. After deduplication and further selection based on the exclusion criteria, an in-depth analysis of each app was performed by two researchers independently. A 71-item six-domain questionnaire about the desirable features of apps was used to assess information, functionalities, and technical features, while the Mobile Application Rating Scale (MARS) was employed for app quality evaluation.

Results: From an initial sample of 684 apps, 22 were deeply analyzed. Most apps did not fulfill the expectations, as just one achieved 50% of all desirable aspects. Postnatal care and counselling for both the mother and child was the least accomplished domain. Moreover, the quality of app information was generally rated more negatively than the quality of their functionality and esthetic features. The lacking aspects were information about methods for postpartum family planning and birth spacing (1/22, 5%) and immunization (2/22, 9%).

Conclusions: The identified gaps could serve as a basis for designing and implementing increasingly high-quality, targeted, and effective apps for pregnancy and postnatal health care, which provide comprehensive, reliable, and evidence-based information, as well as appropriate esthetic and functional characteristics, with relevant implications in terms of maternal and newborn health prevention and promotion.

(J Med Internet Res 2021;23(8):e29151) doi:10.2196/29151

KEYWORDS

pregnancy; postnatal care; app; mHealth; mobile health; newborn

Introduction

Recent development in information and communication technologies has enabled the disruptive expansion of electronic health (eHealth) and mobile health (mHealth). These developments, along with the introduction in clinical practice of technological innovations, such as telemedicine, telemonitoring, and remote screening, are considered essential elements of "game-changing innovations" in the next 25 years [1].

In fact, the widespread distribution of networked devices, which are estimated to reach 29.3 billion in 2023 [2], offers a promising but challenging opportunity of mHealth use for health information seeking, with an important role in health behavior formation [3-5]. In 2017, more than 325,000 health apps were available worldwide [6], and among them, to the best of our knowledge, there were more apps available to support pregnancy than for any other medical domain [7]. These mobile technologies in support of pregnancy have also increased the possibility for both parents and parents-to-be to self-manage health issues, as shown by a recent study conducted in 2019 in Switzerland reporting that 91% of parents declared using digital media for seeking information about their child's health and development [8]. Moreover, a recent meta-analysis showed that social media and mHealth have the potential to be effective in promoting maternal physical health (eg, weight management), mental health, and knowledge about pregnancy [3]. However, when considering apps addressing children's first 1000 days of life, from conception through age 24 months, many of them just focus on the prenatal or postnatal stage [9], failing to consider the continuity between the two phases and their joint impact on maternal and child health.

Furthermore, the reliability, quality, and effectiveness of current available pregnancy and postnatal care apps are unclear, and this could become an obstacle to health promotion, since during pregnancy women are more sensitive to external influences, and misleading information on health care and lifestyle could lead to unnecessary worrying or stress during pregnancy [10].

Considering the large variability among available apps in terms of property, responsibility of information accuracy, level of trustworthiness, and updating of content [11], as well as the lack of a certification system or their classification as a medical device [11,12], the limited evidence of their effectiveness is not a surprise [10]. In addition to this, standard app development and evaluation do not take into account the health literacy level of the target population [12,13], leaving to users the choice of mHealth to use, which therefore may be driven by popularity or esthetic, functional, and engagement aspects. In this regard, the literature suggests that many users do not critically assess the validity of the content provided by apps or consider issues concerning the privacy and security of their personal information and data [14].

Investigating previous research related to the impact of mHealth on maternal and child health care during pregnancy and in the postnatal period, the following two points have come to the forefront: (1) the consideration of children's first 1000 days of life as a crucial developmental window for the children, where it is essential for mothers to receive accurate health information [3] and (2) the critical influence of the health information source on maternal well-being, lifestyle, and decision-making about pregnancy and child health during children's first 1000 days of life [15]. Given these aspects and considering the striking gaps in the current literature, it is an ideal time to conduct an in-depth analysis on pregnancy and postnatal care apps currently in use.

The aim of this review was therefore to critically evaluate available Italian pregnancy and postnatal care apps and to highlight the main shortcomings and resulting future challenges about the apps for the health care of mothers and children in the first 1000 days of life.

Methods

Research on Italian Pregnancy and Postnatal Care Apps

A systematic search was conducted by six independent researchers from Trieste University, IRCCS (Istituto di Ricovero e Cura a Carattere Scientifico) Burlo Garofolo, and Area Science Park for the selection, data extraction, and functional evaluation of the available pregnancy and postnatal care apps on the Apple App Store and Google Play Store. The whole research team included public health specialists, psychologists, sociologists, clinical engineers, and a specialist in ergonomics (all under 45 years of age). The search was conducted for both the Apple App Store and Google Play Store using as keywords the Italian words for *pregnant*, *mother*, *9/nine months*, *birth*, *newborn*, *baby*, *obstetrics*, *pregnancy*, *new baby*, *my baby*, and *child* (the last four keywords were used with their English translation as well). The exclusion criteria for the research were as follows: gaming, photo and video, shopping, and commercial apps; calculator apps; apps specifically developed for health care professionals; fertility or menstrual tracking apps; and apps whose names were not available with Latin alphabets. A basic information set was collected for all resulting apps, including app name, operating system (iOS/Android), free/upon payment availability, availability of further contents upon payment, number of languages, availability of an Italian version, store category, dimension (MB), age restriction, number of downloads, position in download rankings, user rating, first release date, latest update, assistance responsibility, copyright, privacy policy, advertisement, and eventual presence of a medical device European mark (CE). The search was independently conducted between June 15 and July 3, 2020, by six researchers (three working on iOS and three on Android); any inconsistent information collected by the researchers was discussed until consensus was reached. Next, results were merged, and app deduplication was performed.

In-Depth Analysis of Italian Pregnancy and Postnatal Care Apps

Further selection was made among identified apps according to additional exclusion criteria referring to apps not available in the Italian language and in both the Apple App Store and Google Play Store. The apps resulting from this selection were downloaded and installed on both iOS and Android devices for in-depth analysis. The six independent researchers were attributed the same number of apps each. For each app, two independent researchers registered and logged in to evaluate all app contents and functionalities, thus ensuring that each app would be evaluated on both the Apple App Store and Google Play Store. When relevant, a simulation of required input data, such as expected birth date and the starting date of the last menstrual period, was set to fully evaluate app potential. Pursuing this last aim, two evaluators simulated being in the first trimester of pregnancy, another two in the second trimester, and the last two in the third trimester to cover the entire pregnancy period. All investigators were able to consider and evaluate contents related to the postpartum period.

The analysis of the information, functionalities, and technical features of the apps was performed between September 1, 2020, and November 16, 2020, according to a 71-item assessment questionnaire investigating the desirable criteria of apps, developed based on the scientific literature [16], and multiprofessional discussion and agreement (Table 1).

As shown in Table 1, data items referred to six domains corresponding to pregnancy care and counselling, postnatal care and counselling for both the mother and child, reminders and push notifications, notes and records, social support, and app technical features. Each app was evaluated separately for all 71 questions. The answers were attributed as follows: yes (y) if the app provided the information/functionality/feature specified in the question; no (n) if the app did not provide the information/functionality/feature specified in the question; inconsistent information (i) in the case of inconsistent information derived from the operating system (iOS/Android) for the two researchers; and partially (p) if the app only partly provided the information/functionality/feature specified in the question. The Mobile Application Rating Scale (MARS) was used for app quality evaluation regarding four dimensions of objective app quality, including engagement, functionality, esthetics, and information quality. The subjective quality subscale and perceived impact section of the MARS were not included in the evaluation due to possible interresearched biases, as already reported in previous studies [17,18]. All MARS items were rated on a 5-point scale from 1 (inadequate) to 5 (excellent) and required the researcher to circle the number that most accurately represented the quality of the app component under evaluation. The means of the MARS scores judged by the six researchers were calculated for each app. Data were then merged and analyzed using Microsoft Excel for Office 365 ProPlus (Microsoft Corp). Fulfillment of the desirable criteria by each app was expressed as number and percentage on the specific domain and on the total of 71 questions. Responses to the questions by the group of apps was reported as number and percentage with descriptive statistics.

Table 1. Assessment questionnaire.

Domain	Questions
Pregnancy care and counselling	1. Does the app provide information about pregnancy?
	2. Does the app provide information about the woman’s rights during pregnancy (eg, at work, at school, and economical support)?
	3. Are prenatal risks and life-threatening conditions identified in the app?
	4. Does the app inform about maternal physiological and metabolic changes occurring during pregnancy?
	5. Does the app inform about the immunizations that the mother needs to receive?
	6. Does the app provide information about maternal or child services accessibility and contacts?
	7. Does the app provide information about available prenatal diagnostic tests?
	8. Does the app include physical exercises and workouts for women during pregnancy?
	9. Does the app provide pregnancy nutritional counselling for mothers?
	10. Does the app include personal hygiene practices for women during pregnancy?
	11. Does the app provide information about delivery?
	12. Does the app provide information about predelivery courses?
	13. Does the app provide month/trimester-related tips for pregnant women?
	14. Does the app provide a list of essentials for the hospital luggage?
	15. Does the app provide a list of free-of-charge and upon payment examinations during pregnancy?
Postnatal care and counselling for both mother and child	1. Does the app provide a list of essentials for the first welcome of the mother and baby at home?
	2. Does the app inform about maternal physiological and metabolic changes occurring during the postpartum period?
	3. Does the app include information about manifest neonatal complications and warning signs?
	4. Does the app offer information about postpartum mental disorders, such as postpartum depression and baby blues (eg, symptoms and coping strategies)?
	5. Does the app inform about the immunizations that mothers or newborns need during the first 1000 days?
	6. Does the app provide tips for the postpartum recovery process?
	7. Does the app provide practical tips on how to take care of the newborn (eg, hygiene, diapers’ changing, and burping)?
	8. Does the app provide postnatal nutritional counselling for mothers?
	9. Does the app provide a breastfeeding guide and support?
	10. Does the app report personal hygiene practices in the postnatal period?
	11. Does the app encompass methods for postpartum family planning and birth spacing?
Reminders and push notifications	1. Does the app allow to set reminders for medical appointments (eg, prenatal and postnatal check-ups, pediatric visits, and immunizations)?
	2. Does the app include push notification reminders for scheduled medications/immunizations?
	3. Does the app include push notification reminders when the pregnancy month/trimester begins?
	4. Does the app allow to schedule reminders for routine activities (eg, drinking, diapering, feeding, pumping, and sleeping)?
	5. Does the app allow users to change reminders and notifications settings?
Notes and records	1. Does the app require specifying the latest period date or the expected delivery date?
	2. Does the app allow the user to modify the expected delivery date following medical re-evaluation?
	3. Does the app allow to record physiological values of the mother (eg, pressure, temperature, and mood)?
	4. Does the app allow to record contractions or kicks?
	5. Does the app allow to record routine activities of the mother or the newborn (eg, drinking, steps, diapers changes, bottle feeding, and sleeping patterns/times)?
	6. Does the app allow to take note of the medical care the mother or the newborn has received (eg, medications and vaccination shots)?
	7. Does the app allow to track the newborn’s developmental milestones?
	8. Does the app record anthropometric measurements of the fetus (eg, height, weight, and head circumference)?
	9. Does the app record anthropometric measurements of the newborn (eg, height, weight, and head circumference)?
	10. Does the app record measurements of the mother’s weight at baseline and during pregnancy?
	11. Does the app record measurements of the mother’s weight in the postnatal period?
	12. Does the app allow the mother to create a sleep diary for herself?
	13. Does the app allow the mother to create a sleep diary for the newborn?
Social support	1. Is the app integrated with social networks (eg, Facebook and Twitter)?
	2. Is there a FAQ (frequently asked questions) page in the app?
	3. Does the app provide users with social mechanisms to interact with each other and share experiences (eg, community, forum, and chat)?
	4. Does the app provide users with social mechanisms to interact with health care staff (eg, community, forum, and chat)?

Domain	Questions
App technical features	1. Does the app ask users for authentication? 2. Does the app present a privacy policy? 3. If present, is that privacy policy properly written in Italian? 4. Are all app contents freely available to the users (without any payment)? 5. Are there specific inclusion criteria for full app usage (eg, national health service card, place of living, and certification by a health professional)? 6. Does the app require to "sign" an informed consent for app usage? 7. Does the app provide references about the provided contents? 8. Does the app include a glossary of the most used medical terms? 9. Does the app identify the scientific responsibility of the provided contents? 10. Is there a possibility to back-up/restore data within the app? 11. Is there a possibility to download data collected through the app? 12. Does the app have multilanguage support? 13. Does the app geolocalize the user to provide more detailed information? 14. Does the app allow users to book visits, vaccinations, and checkups? 15. Does the app allow users to update their account preferences? 16. Does the app use a tone that is simple, informal, and friendly? 17. Does the app adapt to screen orientation (both portrait and landscape)? 18. Does the app learn user's preferences over time? 19. Does the app implement intuitive and predictable navigation patterns? 20. Are the app contents validated by an institutional source (local, regional, or national)? 21. Is the app a certified medical device according to Italian law? 22. Does the app provide contents through different ways (eg, text, video, and audio)? 23. Does the app ask about user satisfaction?

Results

In total, 684 apps were initially identified; after deduplication, the number reduced to 285. Based on the exclusion criteria, 35 apps were downloaded and installed for in-depth analysis. The flow diagram illustrating the results of the selection process of apps is reported in Figure 1. Following the download of the apps and their further analysis, researchers excluded 13 additional apps as they met at least one of the exclusion criteria or had limitations in content or access that prevented their evaluation. In more detail, one app was not freely available, one was a photo and video app, two were only calculators, one was specifically developed for health professionals, two were menstrual tracking apps (one of which was the only app marked as a medical device among the 35 downloaded and installed apps), two were commercial apps, two contained only physical exercise suggestions, one contained information accessible through a premium account, and one did not work. At the end of the eligibility selection procedure, 22 Italian pregnancy and postnatal care apps underwent in-depth analysis.

There were 180 apps available in the Apple App Store and 148 available in the Google Play Store. Most of these apps (148/180, 82.2% and 146/148, 98.6% for iOS and Android, respectively) were freely available, but many offered extra content upon payment (95/180, 52.8% for iOS and 50/148, 33.8% for Android). Around half of them were available in Italian in both stores, and the majority presented any type of privacy policy. The distributions of the declared presence of advertisements within the app and some age usage restriction were quite different between the two stores. There were significant differences in dimensions among the two as well (Table 2). The mean user rating was well above the average value of 2.5 out of 5 in both stores, being greater in the Google Play Store. Finally, only one app was characterized by the presence of a medical device European mark (CE). The evaluated characteristics are reported in Table 2.

Figure 1. Results of the selection process of apps.

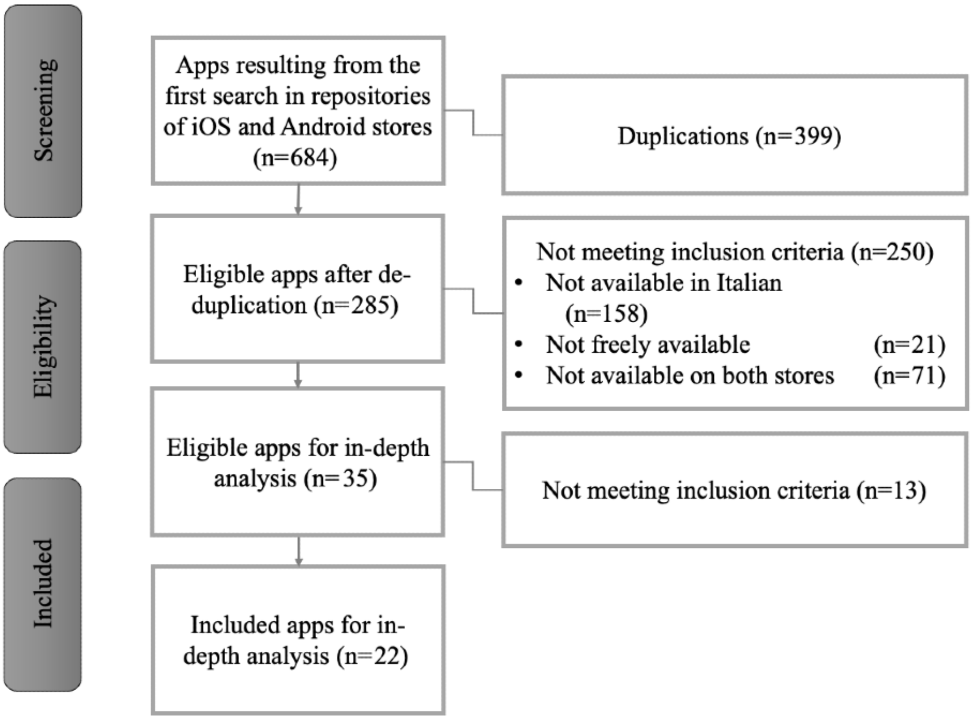

Screening
Eligibility
Included
Apps resulting from the first search in repositories of iOS and Android stores (n=684)
Duplications (n=399)
Eligible apps after de-duplication (n=285)
Not meeting inclusion criteria (n=250)
Not available in Italian (n=158)
Not freely available (n=21)
Not available on both stores (n=71)
Eligible apps for in-depth analysis (n=35)
Not meeting inclusion criteria (n=13)
Included apps for in-depth analysis (n=22)

Table 2. Characteristics of the initial 285 apps for pregnancy and postnatal care identified after deduplication.

App characteristics	Apple App Store (n=180), n (%) or value	Google Play Store (n=148), n (%) or value
Freely available (yes)	148 (82.2%)	146 (98.6%)
Further content upon payment (yes)	95 (52.8%)	50 (33.8%)
Available in Italian (yes)	103 (57.2%)	62 (41.9%)
Privacy policy (yes)	141 (78.3%)	122 (82.4%)
Advertisement (yes)	10 (5.5%)	89 (60.1%)
Mean dimension (MB)	52.3	20.5
Some age restriction (yes)	106 (58.9%)	11 (7.4%)
Mean user rating (stars/5)	3.85	4.22
Medical device European mark (yes)	1	1

In Multimedia Appendix 1, data of in-depth analysis for the 22 evaluated apps are reported. Most apps did not fulfill the expectations, as just one met 50% of all desirable aspects while four met at least 40% of the expected requirements.

As reported in Table 3, the most accomplished domains were pregnancy care and counselling (124/330, 37.6%) and app technical features (160/506, 31.6%), followed by notes and records (77/286, 26.9%), social support (20/88, 22.7%), and postnatal care and counselling for both mother and child (54/242, 22.3%). The least accomplished domain was reminders and push notifications (24/110, 21.8%).

For pregnancy care and counselling, in addition to the general information about pregnancy provided in 14 (64%) cases, other information most frequently given included nutritional counselling for mothers (11/22, 50%), month/trimester-related tips for pregnant women (10/22, 46%), and information about delivery (10/22, 46%). Considering app technical features, a simple, informal, and friendly tone (22/22, 100%), and an intuitive and predictable navigation pattern (19/22, 86%) were the most frequently included features along with the presence of a privacy policy (17/22, 77%). Relating to the various domains, other aspects characterized at least half of the apps, including the presence of a properly written Italian privacy policy (13/22, 59%), as well as a frequently asked questions (FAQ) page (12/22, 55%), the requirement of the latest period date or the expected delivery date (12/22, 55%) and possibility of its eventual modification following medical re-evaluation (12/22, 55%), and the requirement of authentication for app usage (11/22, 50%).

Table 3. Level of accomplishment for each domain.

Domains	Yes[a], n (%)	No[b], n (%)	Inconsistent information[c], n (%)	Partially[d], n (%)
Pregnancy care and counselling (n=330)	124 (37.6%)	188 (57.0%)	3 (0.9%)	15 (4.5%)
Postnatal care and counselling for both mother and child (n=242)	54 (22.3%)	173 (71.5%)	3 (1.2%)	12 (5.0%)
Reminders and push notifications (n=110)	24 (21.8%)	83 (75.5%)	2 (1.8%)	1 (0.9%)
Notes and records (n=286)	77 (26.9%)	200 (69.9%)	0 (0.0%)	9 (3.1%)
Social support (n=88)	20 (22.7%)	67 (76.1%)	0 (0.0%)	1 (1.1%)
App technical features (n=506)[e]	160 (31.6%)	329 (65.0%)	14 (2.8%)	1 (0.2%)

[a] If the app provided the information/functionality/feature.

[b] If the app did not provide the information/functionality/feature.

[c] In the case of inconsistent information derived from the operating system (iOS/Android) of the two researchers.

[d] If the app only partly provided the information/functionality/feature.

[e] Not applicable for two apps.

Aspects that were particularly lacking included information about methods for postpartum family planning and birth spacing (1/22, 5%), immunizations that mothers or newborns need to be administered (2/22, 9%), and postnatal personal hygiene practices for the mother (2/22, 9%). Notably, information about free-of-charge and upon payment examinations during pregnancy was present in only 14% (3/22) of apps, but partial information was given by 32% (7/22) of apps.

Few apps have push notification reminders for scheduled medications/immunizations (3/22, 14%) or when the pregnancy month/trimester begins (3/22, 14%), and a minority of them allow mothers to schedule reminders for routine activities (2/22, 9%) or to take notes of medical care received by both the mother and newborn (2/22, 9%), as well as to monitor the mother's sleep (1/22, 5%). Furthermore, a limited number of apps allow users to record physiological values of the mother (3/22, 14%), routine activities of both the mother and newborn (3/22, 14%), or the newborn's developmental milestones (3/22, 14%). Almost all apps do not provide users with social mechanisms to interact and share experiences with each other (21/22, 96%) or with the health care staff (21/22, 96%). Regarding app technical features, only in a few cases, data collected through the app can be downloaded (3/22, 14%), and a minority of apps have multilanguage support (3/22, 14%) or geolocalize the user to provide more detailed information (3/22, 14%). Just one app requires the habilitation of the pregnant women by the regional health care system as a specific inclusion criterion (1/22, 5%). Only two apps learn user preferences over time (2/22, 9%), but none of them directly allows the mother to book visits, vaccinations, or checkups. Moreover, no app is marked as a medical device according to the medical device directive 93/42/EEC, and subsequent modifications and integrations.

The MARS scores for global app quality ranged from a minimum score of 1.9 to a maximum of 4.5, with most apps (16/22, 73%) achieving a score greater than 3. The engagement score ranged from 1.6 to 4.1 (median 3.0), functionality score ranged from 2.5 to 4.9 (median 4.0), esthetics score ranged from 1.8 to 4.3 (median 3.7), and information score ranged from 1.7 to 4.8 (median 3.5). The mean MARS scores for all 22 analyzed apps are reported in Multimedia Appendix 2.

Discussion

Principal Findings

This systematic search aimed to evaluate the available Italian pregnancy and postnatal care apps, providing an overview of their main characteristics with specific focuses on shortcomings and gaps to be filled with future developments in the eHealth-related field. We believe that such an assessment is indeed crucial for the implementation of increasingly effective apps for pregnancy and postnatal health care. Overall, the presence of a single app containing more than half of the investigated desirable aspects suggests considerable room for improvement in developing apps targeted at the needs of pregnant women and new mothers.

Since during pregnancy women are more sensitive to misleading information [10] and considering the potential of mHealth to be effective in promoting maternal health and knowledge about pregnancy and child health and development [3,8], the importance of developing apps able to provide the most complete, truthful, and reliable information on pregnancy and the postnatal period is unquestionable. In this regard, less than a third of the analyzed apps provided references about included contents or identified their scientific responsibility. Furthermore, the quality of information included in the examined apps was rated more negatively than the quality of their functionality and esthetic features. This aspect can prove to be a double-edged sword, since, as suggested in the literature, the choice of mHealth to be used may be driven by popularity, esthetic features, and functional features rather than the validity, veracity, and reliability of the provided content [14]. Similarly, the use of a simple, informal, and friendly tone, as well as the presence of intuitive and predictable navigation patterns, which are the most frequently included features in the analyzed apps, may pose a threat. While these aspects make an app intuitive, user friendly, and pleasant to use, they may also attract attention and prompt mothers and mothers-to-be to use the app even in the absence of reliable health-related information. In this regard, the tendency to be driven by esthetic or engagement aspects could be particularly detrimental to mothers with low health literacy who lack critical capacity and awareness of the information they receive [12,13]. The development of apps that are both esthetically pleasing and sources of accurate, valid, and comprehensive information, providing indications about the reliability of content, is therefore highly desirable [19].

Moreover, the fact that no app is marked as a medical device denotes a serious shortcoming, which needs to be addressed through the development of apps that are not only information or entertainment tools, but also the subject of a specific regulatory regime, the new medical device regulation. The latter underlines the importance of medical device apps in terms of their predictive utilization for mothers and children, and the control or support of conception.

Regarding provided content, postnatal care and counselling for both the mother and child represented the thematic domain least covered by the analyzed apps, pointing out that most of the available apps focus on pregnancy and neglect the postpartum period. It would therefore be advisable for an app to provide more balanced information on the stages of the mother's life and newborn infant's life, thus paying attention to both prenatal and postnatal care. This point is in line with the literature suggesting that most of the currently available apps just focus on the prenatal or postnatal stage [9], neglecting the continuity between the two phases and their joint impact on the health of both the mother and child.

Surprisingly, information about free-of-charge and upon payment clinical examinations during pregnancy and about immunizations that mothers and newborns need to receive is absent in most apps. This highlights a gap that should be filled through new mobile apps that encourage the promotion of disease prevention for mothers, newborns, and infants and facilitate early intervention in case of problems or complications during pregnancy. An app that provides correct and up-to-date medical information related to preventive or diagnostic medical practices, such as immunizations and examinations, can be a tool to support the health care system by guiding women to the most appropriate screening and therapeutic paths.

Moreover, most of the evaluated apps do not seem to encourage interaction, communication, and confrontation of pregnant women either with other women in the same condition or with health professionals, not taking into consideration the dimension of sociability and the importance of support from both peers and competent professionals. This deficiency, especially in certain periods (eg, during the COVID-19 pandemic), could be burdensome for the woman, potentially generating a sense of loneliness and disorientation.

Very few apps include push notification reminders for scheduled medical appointments, and a minority of apps allow mothers to schedule reminders for routine activities. It follows that the examined apps are not particularly interactive, giving mothers and mothers-to-be few opportunities to rely on them as reference and guidance tools throughout the day or during the pregnancy and postnatal periods. Furthermore, no app directly allows the mother to book visits, vaccinations, or checkups, highlighting the absence of a direct link between the app and reservation systems, and therefore, a lack of integration of this tool within the wider public national or regional health care system. This major gap could be filled by the development of more interconnected and institutional apps that could streamline booking procedures by reducing the number of phone calls or other request modes from patients and the staff involved in managing them.

Furthermore, only two apps could learn user preferences over time, denoting a poor ability of the analyzed apps to adapt to the type of user. Greater "intelligence" of the tool and the consequent possibility of greater customization and profiling of the app system would be desirable, thus resulting in a more tailored option for each woman.

Longitudinal studies, focused on the use of apps by mothers and mothers-to-be over time, could be very useful to highlight which technical features, functionalities, and contents of apps (and to what extent) may be most effective in promoting pregnancy and postnatal health care. The results of such longitudinal research may provide evidence-based guidelines based on which increasingly effective apps could be developed and implemented, thus becoming effective tools for improving health.

Limitations

This study has some limitations, which include only analysis of apps available in Italian and available on both the Apple App Store and Google Play Store without charge. Even if this prevented the researchers from including all apps, this choice was driven by the need to analyze apps with a potentially large user base aiming at equity and equality, and the need for researchers to fully understand the content proposed by the apps. Second, we did not directly involve pregnant women or mothers during the postnatal period as evaluators. Third, even if developed following previous research available in the scientific literature, the definitions of the 71 items for the app evaluation were in part the result of original work of multiprofessional comparison and agreement based on the experience and expertise of the research team. Finally, considering the nature of the dimensions investigated by the MARS, subjectivity in awarding its scores cannot be ruled out. Nevertheless, our research team included members under 45 years of age (mostly women), and some of the women had been pregnant; therefore, we can assume that no biases were there for the assessment procedure. Moreover, in the case of one app, we were prevented from accessing full contents and features, so our evaluation could present some undesired bias in this specific case.

Conclusion

People are increasingly using the internet and specific apps to find health information thanks to the ease of access and the speed of communication that digital solutions offer. While this creates a wide range of advantages and new opportunities, it can pose a threat to public health in terms of prevention and promotion, with exposure to risks, such as lack of content control and misinformation. Within this framework, this study aimed to highlight the main shortcomings in available Italian pregnancy and postnatal care apps to serve as a basis for designing and implementing increasingly high-quality apps that provide comprehensive, reliable, and evidence-based information, and have appropriate esthetic and functional characteristics. Further developments are desirable in the eHealth-related field, with a view to encourage increasingly conscious and effective use of apps by mothers and mothers-to-be for pregnancy and postnatal care.

Acknowledgments

The work has been partially supported by “ARGO system,” a protocol agreement signed in 2018 among the Friuli Venezia Giulia Region, the Italian Ministry of Education University and Research, and the Italian Ministry of Economic Development.

Authors' Contributions

LB designed the research; LB, MDC, FC, MC, MP, FS, and CDV collected data; LB, MDC, FC, MC, MP, FS, AS, CDV, MB, and SZ discussed the investigation methodology and contributed to result interpretation; LB and CDV performed data analysis; LB and SZ supervised the study conduction; LB and CDV wrote the original draft; and TS revised contents. All authors revised the paper and agreed with the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Total number and percentage of accomplishments for each questionnaire item and globally for each of the 22 apps.

[DOCX File , 26 KB - jmir_v23i8e29151_app1.docx]

Multimedia Appendix 2

Mobile Application Rating Scale scores for the 22 evaluated apps.

[DOCX File , 17 KB - jmir_v23i8e29151_app2.docx]

References

1. Engmann CM, Khan S, Moyer CA, Coffey PS, Bhutta ZA. Transformative Innovations in Reproductive, Maternal, Newborn, and Child Health over the Next 20 Years. PLoS Med 2016 Mar 2;13(3):e1001969 [FREE Full text] [doi: 10.1371/journal.pmed.1001969] [Medline: 26933951]
2. Cisco Annual Internet Report (2018–2023) White Paper. Cisco. 2020. URL: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html [accessed 2021-07-26]
3. Chan KL, Chen M. Effects of Social Media and Mobile Health Apps on Pregnancy Care: Meta-Analysis. JMIR Mhealth Uhealth 2019 Jan 30;7(1):e11836 [FREE Full text] [doi: 10.2196/11836] [Medline: 30698533]
4. Guerra-Reyes L, Christie VM, Prabhakar A, Harris AL, Siek KA. Postpartum Health Information Seeking Using Mobile Phones: Experiences of Low-Income Mothers. Matern Child Health J 2016 Nov 17;20(Suppl 1):13-21 [FREE Full text] [doi: 10.1007/s10995-016-2185-8] [Medline: 27639571]
5. Lee Y, Moon M. Utilization and Content Evaluation of Mobile Applications for Pregnancy, Birth, and Child Care. Healthc Inform Res 2016 Apr;22(2):73-80 [FREE Full text] [doi: 10.4258/hir.2016.22.2.73] [Medline: 27200216]
6. mHealth App Economics 2017: Current status and future trends in mobile health. Research 2 Guidance. 2017. URL: https://research2guidance.com/wp-content/uploads/2017/10/1-mHealth-Status-And-Trends-Reports.pdf [accessed 2021-07-26]
7. Tripp N, Hainey K, Liu A, Poulton A, Peek M, Kim J, et al. An emerging model of maternity care: smartphone, midwife, doctor? Women Birth 2014 Mar;27(1):64-67. [doi: 10.1016/j.wombi.2013.11.001] [Medline: 24295598]
8. Jaks R, Baumann I, Juvalta S, Dratva J. Parental digital health information seeking behavior in Switzerland: a cross-sectional study. BMC Public Health 2019 Feb 21;19(1):225 [FREE Full text] [doi: 10.1186/s12889-019-6524-8] [Medline: 30791927]
9. Zhang P, Dong L, Chen H, Chai Y, Liu J. The Rise and Need for Mobile Apps for Maternal and Child Health Care in China: Survey Based on App Markets. JMIR Mhealth Uhealth 2018 Jun 08;6(6):e140 [FREE Full text] [doi: 10.2196/mhealth.9302] [Medline: 29884601]
10. Overdijkink SB, Velu AV, Rosman AN, van Beukering MD, Kok M, Steegers-Theunissen RP. The Usability and Effectiveness of Mobile Health Technology-Based Lifestyle and Medical Intervention Apps Supporting Health Care During Pregnancy: Systematic Review. JMIR Mhealth Uhealth 2018 Apr 24;6(4):e109 [FREE Full text] [doi: 10.2196/mhealth.8834] [Medline: 29691216]
11. Biviji R, Vest JR, Dixon BE, Cullen T, Harle CA. Factors Related to User Ratings and User Downloads of Mobile Apps for Maternal and Infant Health: Cross-Sectional Study. JMIR Mhealth Uhealth 2020 Jan 24;8(1):e15663 [FREE Full text] [doi: 10.2196/15663] [Medline: 32012107]
12. Taki S, Campbell KJ, Russell CG, Elliott R, Laws R, Denney-Wilson E. Infant Feeding Websites and Apps: A Systematic Assessment of Quality and Content. Interact J Med Res 2015 Sep 29;4(3):e18 [FREE Full text] [doi: 10.2196/ijmr.4323] [Medline: 26420339]

13. Chen Q, Carbone ET. Functionality, Implementation, Impact, and the Role of Health Literacy in Mobile Phone Apps for Gestational Diabetes: Scoping Review. JMIR Diabetes 2017 Oct 04;2(2):e25 [FREE Full text] [doi: 10.2196/diabetes.8045] [Medline: 30291088]
14. Lupton D, Pedersen S. An Australian survey of women's use of pregnancy and parenting apps. Women Birth 2016 Aug;29(4):368-375. [doi: 10.1016/j.wombi.2016.01.008] [Medline: 26874938]
15. Criss S, Woo Baidal JA, Goldman RE, Perkins M, Cunningham C, Taveras EM. The Role of Health Information Sources in Decision-Making Among Hispanic Mothers During Their Children's First 1000 Days of Life. Matern Child Health J 2015 Nov 30;19(11):2536-2543 [FREE Full text] [doi: 10.1007/s10995-015-1774-2] [Medline: 26122256]
16. Sardi L, Idri A, Redman LM, Alami H, Bezad R, Fernández-Alemán JL. Mobile health applications for postnatal care: Review and analysis of functionalities and technical features. Comput Methods Programs Biomed 2020 Feb;184:105114 [FREE Full text] [doi: 10.1016/j.cmpb.2019.105114] [Medline: 31655305]
17. Knitza J, Tascilar K, Messner E, Meyer M, Vossen D, Pulla A, et al. German Mobile Apps in Rheumatology: Review and Analysis Using the Mobile Application Rating Scale (MARS). JMIR Mhealth Uhealth 2019 Aug 05;7(8):e14991 [FREE Full text] [doi: 10.2196/14991] [Medline: 31381501]
18. 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]
19. Bert F, Passi S, Scaioli G, Gualano MR, Siliquini R. There comes a baby! What should I do? Smartphones' pregnancy-related applications: A web-based overview. Health Informatics J 2016 Sep 26;22(3):608-617 [FREE Full text] [doi: 10.1177/1460458215574120] [Medline: 25900813]

Abbreviations

MARS: Mobile Application Rating Scale

Edited by R Kukafka; submitted 01.04.21; peer-reviewed by R Krukowski, M Reynolds; comments to author 23.04.21; revised version received 10.05.21; accepted 24.05.21; published 10.08.21.

Please cite as:
Brunelli L, De Vita C, Cenedese F, Cinello M, Paris M, Samogizio F, Starec A, Bava M, Dal Cin M, Zanchiello S, Stampalija T
Gaps and Future Challenges of Italian Apps for Pregnancy and Postnatal Care: Systematic Search on App Stores
J Med Internet Res 2021;23(8):e29151
URL: https://www.jmir.org/2021/8/e29151
doi:10.2196/29151
PMID:34383668

©Laura Brunelli, Chiara De Vita, Fabrizio Cenedese, Michela Cinello, Marta Paris, Francesca Samogizio, Anja Starec, Michele Bava, Margherita Dal Cin, Sara Zanchiello, Tamara Stampalija. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Feasibility of an Interactive Health Coaching Mobile App to Prevent Malnutrition and Muscle Loss in Esophageal Cancer Patients Receiving Neoadjuvant Concurrent Chemoradiotherapy: Prospective Pilot Study

Kyungmi Yang[1], MD, PhD; Dongryul Oh[1], MD, PhD; Jae Myoung Noh[1], MD, PhD; Han Gyul Yoon[1], MD; Jong-Mu Sun[2], MD, PhD; Hong Kwan Kim[3], MD, PhD; Jae Ill Zo[3], MD, PhD; Young Mog Shim[3], MD, PhD; Hyunyoung Ko[4], MSc; Jungeun Lee[4,5], PhD; Youngin Kim[6,7], MD

[1]Department of Radiation Oncology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea

[2]Department of Medicine, Division of Hematology-Oncology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea

[3]Department of Thoracic and Cardiovascular Surgery, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea

[4]Noom Korea Inc, Seoul, Republic of Korea

[5]Ingenium College of Liberal Arts, KwangWoon University, Seoul, Republic of Korea

[6]Noom Inc, New York, NY, United States

[7]Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, Republic of Korea

Corresponding Author:
Dongryul Oh, MD, PhD
Department of Radiation Oncology
Samsung Medical Center
Sungkyunkwan University School of Medicine
81 Irwon-ro
Gangnam-gu
Seoul, 06351
Republic of Korea
Phone: 82 2 3410 2612
Fax: 82 2 3410 2619
Email: dongryul.oh@samsung.com

Abstract

Background: Excessive muscle loss is an important prognostic factor in esophageal cancer patients undergoing neoadjuvant chemoradiotherapy (NACRT), as reported in our previous research.

Objective: In this pilot study, we prospectively tested the feasibility of a health coaching mobile app for preventing malnutrition and muscle loss in this patient population.

Methods: Between July 2019 and May 2020, we enrolled 38 male patients with esophageal cancer scheduled for NACRT. For 8 weeks from the start of radiotherapy (RT), the patients used Noom, a health coaching mobile app that interactively provided online advice about food intake, exercise, and weight changes. The skeletal muscle index (SMI) measured based on computed tomography and nutrition-related laboratory markers were assessed before and after RT. We evaluated the changes in the SMI, nutrition, and inflammatory factors between the patient group that used the mobile app (mHealth group) and our previous study cohort (usual care group). Additionally, we analyzed the factors associated with walk steps recorded in the app.

Results: Two patients dropped out of the study (no app usage; treatment changed to a definitive aim). The use (or activation) of the app was noted in approximately 70% (25/36) of the patients until the end of the trial. Compared to the 1:2 matched usual care group by propensity scores balanced with their age, primary tumor location, tumor stage, pre-RT BMI, and pre-RT SMI level, 30 operable patients showed less aggravation of the prognostic nutritional index (PNI) (–6.7 vs –9.8; *P*=.04). However, there was no significant difference in the SMI change or the number of patients with excessive muscle loss (ΔSMI/50 days >10%). In patients with excessive muscle loss, the walk steps significantly decreased in the last 4 weeks compared to those in the first 4 weeks. Age affected the absolute number of walk steps (*P*=.01), whereas pre-RT sarcopenia was related to the recovery of the reduced walk steps (*P*=.03).

Conclusions: For esophageal cancer patients receiving NACRT, a health care mobile app helped nutritional self-care with less decrease in the PNI, although it did not prevent excessive muscle loss. An individualized care model with proper exercise as well as nutritional support may be required to reduce muscle loss and malnutrition.

(J Med Internet Res 2021;23(8):e28695) doi:10.2196/28695

KEYWORDS

esophageal cancer; malnutrition; muscle loss; sarcopenia; mobile app; mHealth

Introduction

Esophageal cancer is one of the most aggressive malignancies and is ranked as the sixth leading cause of cancer-related death worldwide [1]. Recent advances in treatment including radiotherapy (RT), chemotherapy, and surgery, and their combinations have led to improved clinical outcomes; however, patients continue to experience high mortality [2]. Patients with esophageal cancer commonly have symptoms such as dysphagia and weight loss even before a confirmed diagnosis of the disease, and up to 80% of patients are nutritionally compromised [3,4]. Malnutrition, cachexia, and sarcopenia are reported as poor prognostic factors associated with treatment compliance and clinical outcomes in esophageal cancer and various cancers other as well [5,6].

Previously, we conducted a retrospective review among 248 esophageal cancer patients who underwent surgery and reported excessive muscle loss after neoadjuvant concurrent chemoradiotheapy (NACRT) as a significant poor prognostic factor for disease recurrence and overall survival [7]. In that study, many patients experienced malnutrition before and after NACRT; 62.9% of the patients were assigned the status of sarcopenia before the start of the treatment, and 28.2% showed more than 10% deterioration in the skeletal muscle index (SMI) after NACRT. To reduce muscle loss, more active support toward the patient's nutrition and physical activity was needed than education from a physician or nutrition specialist. Additionally, most patients were managed in the outpatient setting.

Noom (Noom Inc) [8] is one of the most popular commercial health care mobile apps. This app has been mainly used by obese people for weight control [9,10]. In addition to losing weight, the company also provides paramedical experience for patients with diabetes or eating disorders, and pregnant women [11-14]. It provides interactive health coaching on various aspects of nutrition, exercise, and weight control. Although there is little evidence for the usage of this app, it can be applied to the self-management of nutrition for cancer patients.

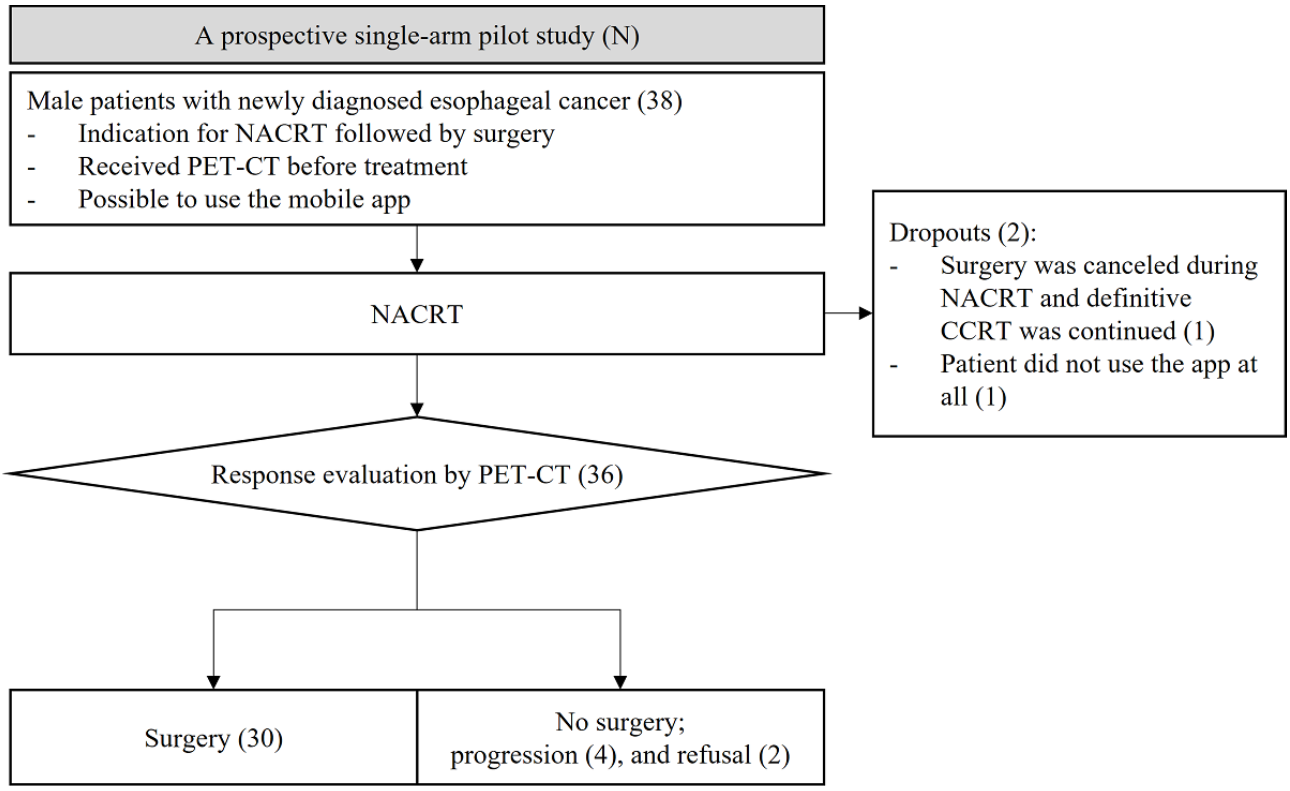
This prospective pilot study aimed to evaluate the usefulness of a health care mobile app in preventing malnutrition and excessive muscle loss in patients with esophageal cancer receiving NACRT.

Methods

Study Design

The inclusion criteria for this study were as follows: adult male patients (1) newly diagnosed with esophageal cancer, (2) scheduled for NACRT followed by radical surgery as per the authors' institutional protocol, and (3) undergoing positron emission tomography-computed tomography (PET-CT) as part of the diagnostic workup (Figure 1). Patients with synchronous distant metastasis, previous history of thoracic irradiation, or having difficulty using an app on a smartphone were excluded. Between July 2019 and May 2020, totally 38 patients were enrolled in this study. Data regarding the app usage were gathered for 8 weeks from the beginning of NACRT. During the trial, two patients dropped out; one did not use the app at all, and the other decided not to undergo surgery during NACRT and continued RT with a definitive aim for 6 weeks. Finally, 36 patients analyzed to evaluate the feasibility and effectiveness of using the app. This study was approved by the Institutional Review Board of Samsung Medical Center, and all participants provided written informed consent.

Figure 1. Flowchart of the study design. CCRT: concomitant radiotherapy and chemotherapy; NACRT: neoadjuvant chemoradiotherapy; PET-CT: positron emission tomography-computed tomography.


A prospective single-arm pilot study (N)
Male patients with newly diagnosed esophageal cancer (38)
- Indication for NACRT followed by surgery
- Received PET-CT before treatment
- Possible to use the mobile app
NACRT
Dropouts (2):
- Surgery was canceled during NACRT and definitive CCRT was continued (1)
- Patient did not use the app at all (1)
Response evaluation by PET-CT (36)
Surgery (30)
No surgery; progression (4), and refusal (2)

Treatment Scheme

Details of the treatment scheme, which includes RT and chemotherapy for NACRT, are described in our previous research [7]. The dose prescriptions for RT were 44 Gy in 22 fractions (2.0 Gy per fraction) for 18 patients and 43 Gy in 20 fractions (2.15 Gy per fraction) for 18 patients. All patients underwent intensity-modulated RT. All the patients completed NACRT as scheduled without grade 4 or higher complications. The concurrent chemotherapeutic regimen involved 2 cycles of cisplatin-based therapy in combination with 5-fluorouracil or capecitabine during RT. For response evaluation of NACRT, PET-CT was performed 3-4 weeks after the completion of NACRT, before surgery. The median time interval between pre-RT and post-RT PET-CTs was 73 days, ranging from 54 to 103 days. Patients underwent radical thoracic surgery, and the mean interval between the beginning of RT and completion of surgery was 10.5 weeks (range 9.1-14 weeks). Finally, surgery was not performed in 6 patients; 4 patients showed progressive disease after NACRT, not to assure curative surgery, and 2 patients did not want to undergo surgery after NACRT. Before NACRT, prophylactic feeding tube insertion was not performed in all patients except 1. During the trial, all patients received the best supportive care for their symptoms, such as painkillers and antacid drugs if needed.

Mobile App for Health Coaching

Before the start of NACRT, patients were informed about the trial and app usage by a study instructor and attended an online meeting with the app manager. The app provided services following enrollment, and the patients used it autonomously. The app setting was similar to that used in previous studies [9]. During the first sign-in, patients set their information including age, sex, height, and current and target weight. Users could record their food intake, exercise, and weight on a daily basis. In addition, the activity monitor in the app automatically measured the number of walk steps. Based on these records, the app provided summaries on calorie balance and weight trends. Additionally, experts in nutrition and exercise therapy provided feedback messages and encouraged the users through the app. The number of active records and messages from patients was counted each week in terms of nutrition, weight, and exercise. If any active data were recorded by the patient in a certain week, then the app was considered “activated” for that week. Based on previous experiences of the app users, as the compliance for this trial, the activation level of the app per 8 weeks was categorized as follows: “high,” activated for 6-8 weeks; “moderate,” activated for 4-5 weeks; and “low,” activated for 1-3 weeks.

Muscle Loss and Malnutrition Assessment

Similar to our previous study, the skeletal muscle area (cm^2) at the level of the third lumbar vertebra (L3) in the two CT image sets from PET-CT before and after NACRT was measured using an in-house software program [15]. The SMI (cm^2/m^2) was calculated from the skeletal muscle area divided by the square of the height (m^2). We also used the cutoff value of 52.4 cm^2/m^2 for sarcopenia from our previous study [7] and a population-based study [16]. ΔSMI (%), which is the percentage of SMI change based on the pre-RT SMI, was adjusted with various time intervals between 2 PET-CT measurements, and unified in 50 days per interval between 2 PET-CT measurements for each patient (ΔSMI/50 days, %). Excessive muscle loss was defined when ΔSMI/50 days was more than 10%.

Other information regarding nutritional status was obtained from the laboratory tests on the day closest to that before the start of RT (pre-RT) and those as preoperative workups (post-RT), which were the first blood tests in most patients performed after the end of RT. Data were collected for the following laboratory parameters: absolute counts of white blood cells (WBC, /µL), absolute neutrophil count (ANC, /µL), absolute lymphocyte count (ALC, /µL), platelet count (/µL), and albumin level (g/dL). From these results, the neutrophil-to-lymphocyte ratio (NLR, ANC/ALC), platelet-to-lymphocyte ratio (platelet/ALC), and prognostic nutritional index (PNI, 10 × albumin + 0.005 × ALC) [17] were calculated.

Statistical Analyses

The purpose of this study was to evaluate whether ΔSMI (%/50 days) was reduced in patients who used the mobile app (mHealth group) compared with that in our previous study cohort (usual care group). Additionally, changes in nutritional and inflammatory factors were also analyzed between the two groups. To compare the two groups, propensity score matching was used to balance characteristics such as the age, primary tumor location, tumor stage, pre-RT BMI, and pre-RT SMI level. Fisher tests for discrete variables and *t* tests (or Mann-Whitney *U* test) for continuous variables were conducted for comparing the characteristics. To identify significant factors associated with excessive muscle loss in this study, logistic regression was performed with the quantified app data. Additionally, we performed mixed model analysis for repeated measures of walk steps counted by the app for 8 weeks. R 4.0.3 (version 4.0.3, R Development Core Team) [18] and SPSS Statistics (version 27.0, IBM Corp) were used for statistical analyses with *P*<.05 suggesting statistical significance.

Results

Compliance With the App

For 8 weeks of the trial, 80.6% (29/36), 77.8% (28/36), 75.0% (27/36), 77.8% (28/36), 72.2% (26/36), 66.7% (24/36), and 72.2% (26/36) of the patients had been active on the app from the second week to the 8th week. In addition, the activation levels were as follows: high, moderate, and low levels in 25 (69.4%), 3 (8.3%), and 8 (22.2%) patients, respectively. However, the patients selectively used parts of the app (Table 1). The most activated records were for the nutritional part. However, in the exercise records, only one was moderately activated, and there was no highly activated patient. Walk steps were passively gathered and had a pattern similar to that of the total activation level.

Table 1. Activation level based on the number of activated weeks.

Activated part of the app	Activation level[a], n (%) (N=36)		
	Low	Moderate	High
At least one activation in any part of the app	8 (22.2)	3 (8.3)	25 (69.4)
Nutrition	14 (38.8)	7 (19.4)	15 (41.7)
Weight	20 55.6)	4 (11.1)	12 (33.3)
Exercise	35 (97.2)	1 (2.8)	0 (0)
Walking[b]	6 (16.7)	5 (13.9)	25 (69.4)

[a] Activation level was determined depending on the number of activated weeks—low: activated for 1 to 3 weeks; moderate: activated for 4 to 5 weeks; high: activated for 6 to 8 weeks.

[b] The number of walk steps was automatically recorded by the activated app.

Patient Characteristics and Comparison With the Usual Care Group

The median patient age was 59 years, and all patients had an Eastern Cooperative Oncology Group Performance Status (ECOG PS) of 0 to 1. The details of the patient characteristics are summarized in Table 2. Compared to the previous study cohort (usual care group), the age, tumor location, and some of the laboratory markers were significantly different in this study cohort (mHealth group) (Supplement table). As a result, ΔSMI/50 days (%) and the proportion of patients with excessive muscle loss (ΔSMI/50 days >10%) were not significantly different between the two groups. However, in the mHealth group, the PNI decreased to a lesser extent after NACRT.

After 1:2 propensity score matching to adjust for variables such as the age, primary tumor location, tumor stage, pre-RT BMI, and pre-RT SMI level, we compared 30 patients in the mHealth group with those of the usual care group (Table 2). Even after propensity score matching, ΔSMI/50 days (%) and the proportion of patients with excessive muscle loss (ΔSMI/50 days >10%) were not significantly different (–8.1% vs 7.4%, *P*=.57 and 33.3% vs 30.0%, *P*=.94, respectively). The PNI decreased to a lesser extent in the mHealth group than in the usual care group (–6.7 vs –9.8, *P*=.04).

Table 2. Propensity score matching and comparison in 30 patients who underwent surgery after neoadjuvant concurrent chemoradiotherapy.

Characteristics	Previous cohort [7] (n=60)	mHealth group (n=30)[a]	*P* value
Age (years), median (SD)	58.5 (7.8)	59.2 (6.5)	.8
Age group, n (%)			.99
<60 years	34 (56.7)	17 (56.7)	
≥60 years	26 (43.3)	13 (43.3)	
Smoking status, n (%)			.41
Current smoker	35 (58.3)	14 (46.7)	
Ex- or nonsmoker	25 (41.7)	16 (53.3)	
Tumor location, n (%)			.73
Upper	12 (20)	4 (13.3)	
Middle	21 (35)	11 (36.7)	
Lower	27 (45)	15 (50)	
cT stage[b], n (%)			.49
1-2	24 (40)	9 (30)	
3-4	36 (60)	21 (70)	
cN stage[c], n (%)			.94
0-1	34 (56.7)	18 (60)	
2-3	26 (43.3)	12 (40)	
Pre-RT[d] BMI (kg/m^2), mean (SD)	22.9 (2.8)	22.9 (2.3)	.99
Pre-RT BMI (kg/m^2)			.99
<20, n (%)	7 (11.7)	3 (10)	
≥20, n (%)	53 (88.3)	27 (90)	
Post-RT BMI (kg/m^2), mean (SD)	22 (6)	21.8 (2.6)	.65
Post-RT BMI (kg/m^2)			
<20, n (%)	10 (16.7)	7 (23.3)	.63
≥20, n (%)	50 (83.3)	23 (76.7)	
ΔBMI (kg/m^2), mean (SD)	-0.8 (1.4)	-1.1 (1.2)	.38
Pre-RT SMI[e] (cm^2/m^2), mean (SD)	51.0 (9.1)	51.3 (6.2)	.87
Pre-RT SMI (cm^2/m^2)			.99
Sarcopenia, n (%)	36 (60)	18 (60)	
Nonsarcopenia, n (%)	24 (40)	12 (40)	
Post-RT SMI (cm^2/m^2), mean (SD)	45 (7.8)	45.9 (6.9)	.58
Post-RT SMI (cm^2/m^2)			.93
Sarcopenia, n (%)	48 (80)	23 (76.7)	
Nonsarcopenia, n (%)	12 (20)	7 (23.3)	
ΔSMI (/50 days), mean, % (SD)	–8.1 (5.3)	–7.4 (6.5)	.57
ΔSMI (/50 days)			.94
Decreased >10%, n (%)	20 (33.3)	9 (30)	
Decreased <10%, n (%)	40 (66.7)	21 (70)	
Laboratory markers			

Characteristics	Previous cohort [7] (n=60)	mHealth group (n=30)[a]	*P* value
WBC[f] (×10^3/μL), mean (SD)			
Pre-RT	8.2 (2.5)	7.1 (1.7)	.02
Post-RT	4.8 (2.2)	4.7 (2.8)	.93
Δ	–3.4 (3.5)	-2.3 (2.8)	.05
ANC[g] (×10^3/μL), mean (SD)			
Pre-RT	5.1 (2.2)	4.4 (1.3)	.2
Post-RT	2.8 (2)	2.6 (1.6)	.6
Δ	–2.4 (3.1)	–1.8 (2.1)	.34
ALC[h] (×10^3/μL), mean (SD)			
Pre-RT	2.2 (0.5)	1.9 (0.6)	.02
Post-RT	1.3 (0.6)	1.5 (1.3)	.6
Δ	–0.9 (0.7)	–0.4 (1.3)	.11
Platelet(×10^3/μL), mean (SD)			
Pre-RT	263.0 (72.2)	260.3 (52.9)	.86
Post-RT	215.0 (70.3)	203.5 (74)	.48
Δ	–48.0 (69.3)	–56.7 (84.2)	.6
Albumin (g/dL), mean (SD)			
Pre-RT	4.4 (0.3)	4.4 (0.3)	.19
Post-RT	3.8 (0.5)	4.0 (0.4)	.07
Δ	–0.6 (0.4)	–0.5 (0.4)	.34
NLR[i], >mean (SD)			
Pre-RT	2.4 (1.0)	2.5 (1)	.77
Post-RT	2.8 (3.7)	3.1 (4.7)	.75
Δ	0.4 (3.9)	0.6 (5.1)	.82
PLR[j], mean (SD)			
Pre-RT	125.3 (41.8)	149.8 (54.1)	.02
Post-RT	209.4 (159.6)	212.1 (160.8)	.94
Δ	84.1 (157.6)	62.4 (173.4)	.55
PNI[k], mean (SD)			
Pre-RT	54.6 (4.2)	54 (4.2)	.54
Post-RT	44.8 (5.7)	47.3 (7.7)	.09
Δ	–9.8 (6)	–6.7 (7.5)	.04

[a] Patients who received surgery.

[b] cT: clinical tumor stage.

[c] cN: clinical nodal stage.

[d] RT: radiotherapy.

[e] SMI: skeletal muscle index.

[f] WBC: white blood cells.

[g] ANC: absolute neutrophil count.

[h] ALC: absolute lymphocyte count.

[i] NLR: neutrophil-to-lymphocyte ratio.

[j] PLR: platelet-to-lymphocyte ratio.

[k] PNI: prognostic nutritional index.

Use of the App and the Associated Excessive Muscle Loss

In this study, there was excessive muscle loss in 12 of the 36 patients (33.3%) (Table 3). Regarding the app usage, there was no significant difference in the activation level, but walk steps decreased more in the excessive muscle loss group. More patients with a 70% or more decrease in walk steps between the first and the last 4 weeks were in the excessive muscle loss group (66.7% (6/9) vs 14.3% (3/21), *P*=.02).

Table 3. App usage and excessive muscle loss.

App usage	Nonexcessive muscle loss group (n=24)	Excessive muscle loss group (n=12)	*P* value
App activation level, n (%)			.96
Low	5 (20.8)	3 (25)	
Moderate	2 (8.3)	1 (8.3)	
High	17 (70.8)	8 (66.7)	
First to fourth week, n (%)			.99
No full activation	6 (25)	3 (25)	
Full activation	18 (75)	9 (75)	
Fifth to eighth week, n (%)			.11
No full activation	6 (25)	7 (58.3)	
Full activation	18 (75)	5 (41.7)	
Walk steps[a]			
Steps per week/10^3, mean (SD)	15.8 (13.1)	16.8 (17)	.85
First to fourth week (a), mean (SD)	16.9 (13.2)	18.4 (18.2)	.79
Fifth to eighth week (b), mean (SD)	10.4 (10.5)	5.6 (9)	.22
Δ (b–a, %), mean (SD)	–33.1 (64.4)	–69.0 (35.5)	.13
Decrease <70%, n (%)	18 (85.7)	3 (33.3)	.02
Decrease >70%, n (%)	3 (14.3)	6 (66.7)	—[b]

[a] Nonexcessive muscle loss group (n=21); excessive muscle loss group (n=9).

[b] Not applicable.

Mixed Model Analysis for Weekly Walk Steps

Repeatedly measured walk steps appeared to have a pattern (Figure 2A), which decreased at the 5th and 6th weeks and slightly recovered at the 7th and 8th weeks. Discrete variables such as age, BMI, SMI, and surgery were used for this analysis (Table 4). When the variables were fixed, time (week) was commonly significant. In contrast, when time was fixed, only the age was significantly different (*P*=.01), which showed a gap in the absolute number of walk steps in the two age groups (Figure 2B). For the interaction between time and each variable, only pre-RT sarcopenia was significantly different (*P*=.03), which showed different recovery patterns at the 6th and 8th weeks between the pre-RT sarcopenia and nonsarcopenia groups (Figure 2C). Although not significant, excessive muscle loss seemed to affect the trend in the number of walk steps (Figure 2D).

Figure 2. Weekly trends in walk steps recorded automatically by the application for 8 weeks (N=33): all patients (A), patient groups according to age (B), pre-RT sarcopenia (C), and excessive muscle loss (-ΔSMI/50 days >10%), during neoadjuvant concurrent chemoradiotherapy. RT: radiotherapy; SMI: skeletal muscle index.

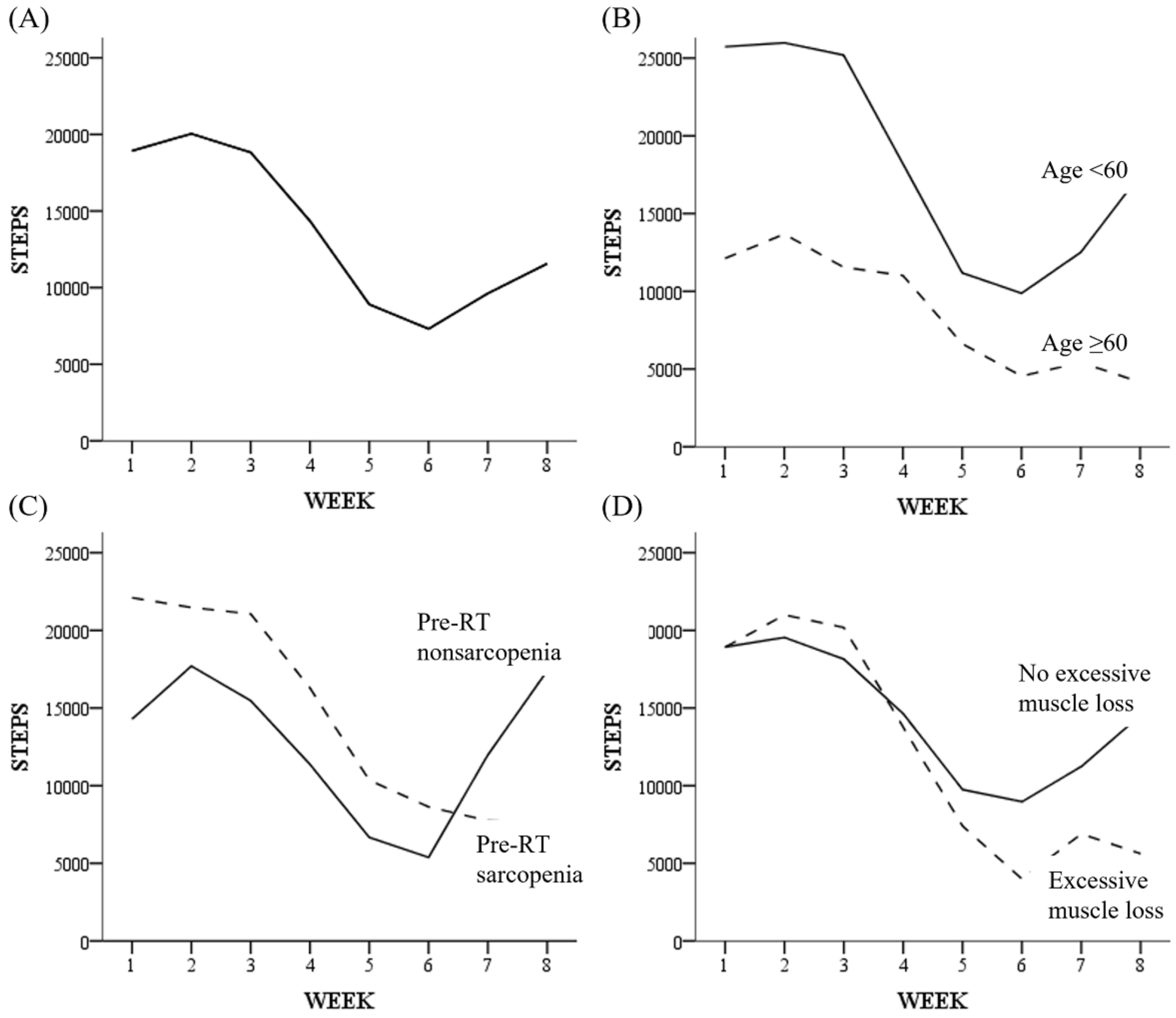


Table 4. Mixed model analysis with walk steps for 8 weeks.

For 8 weeks	*P* value for weeks	*P* value for variables	*P* value for interaction
Age (<60 vs ≥60 years)	<.001	*.006*	.39
Pre-RT[a] BMI (≥20 vs <20 kg/m^2)	.001	.17	.98
Post-RT BMI (≥20 vs <20 kg/m^2)	.001	.67	.47
-ΔBMI (>5% vs <5% kg/m^2)	<.001	.98	.096
Pre-RT sarcopenia (yes vs no)	<.001	.55	.03
Post-RT sarcopenia (yes vs no)	.001	.44	.64
-ΔSMI[b]/50 days >10% (>10% vs <10%)	<.001	.42	.43
Surgery (Yes vs No)	.007	.08	.96

[a] RT: radiotherapy.

[b] SMI: skeletal muscle index.

Discussion

Principal Findings

Sarcopenia is a syndrome characterized by a progressive decrease in skeletal muscle mass and strength, leading to a risk of worse outcomes in terms of physical ability, quality of life, and survival, although there are various definitions for this term provided by the European Working Group on Sarcopenia in Older People (EWGSOP), the European Society for Clinical Nutrition and Metabolism Special Interest Groups (ESPEN-SIG), and the International Working Group on Sarcopenia (IWGS) [19]. It was first considered an age-related disease by Rosenberg [20]. However, in recent years, it is clinically noticeable that sarcopenia can be caused by malignancy or inflammatory diseases and adversely affects their treatment. The mechanism of sarcopenia is not clear but seems multifactorial; its risk factors are old age; female sex; low level of physical activity; and comorbidities including malignancy, obesity, osteoporosis, and diabetes [21]. Focusing on malignancy, higher metabolism, and inflammation by aggressive cancer cells, cancer treatments including surgery, chemotherapy, and RT; anorexia or poor oral intake; and low physical activity make patients more susceptible to sarcopenia [22]. Moreover, gastrointestinal malignancy easily causes malnutrition, and most esophageal cancer patients experience cachexia or sarcopenia. Furthermore, the presence of sarcopenia in cancer patients has been accepted clinically as one of the important predictors of survival and treatment outcomes [6,22]. In addition to clinical evidence, the muscle loss is gaining interest in biochemical research, including the anti-inflammatory or anticancer effects of myokines [23]. We previously analyzed the effect of sarcopenia in patients with esophageal cancer receiving NACRT [7]. NACRT itself has risk factors for sarcopenia; symptoms or complications such as nausea/vomiting and anorexia induced by chemotherapy and acute esophagitis by RT can deplete nutrition and physical energy [22]. We revealed that the changes in the muscle index (ΔSMI/50 days) were the prognostic factors for disease recurrence and survival and not the presence of sarcopenia before the treatment [7]. Additionally, sarcopenia is related to nutritional and inflammatory markers [7,24]. Therefore, we suggest that appropriate intervention for nutrition and physical activity could be beneficial for these patients.

Numerous interventions to overcome malnutrition in patients with various cancers have been attempted [3,25,26]. Unfortunately, interventions to improve cancer patients' malnutrition or muscle loss have been rare. Malnutrition is still considered insurmountable, and most esophageal cancer patients find it difficult to overcome malnutrition or often experience worsened conditions. This study, which used an interactive health coaching mobile app as an intervention for nutrition and physical activity, found similar negative results in terms of preventing muscle loss in esophageal cancer patients receiving NACRT compared to that in the usual care group. As a secondary endpoint, however, the use of the mobile app could improve nutritional indicators such as PNI, which is also known to be a prognostic factor for gastrointestinal cancers [17]. The compliance of the mobile app was as good because 70% (20/36) of the patients used it until the end of the trial, especially in nutrition and weight items. This could be because the patients were concerned about maintaining their oral intake and weight.

Guidelines for nutritional support suggest that appropriate interventions have to be selected depending on each patient's nutritional status, from basic nutritional modification and oral nutritional supplements to enteral or parenteral nutrition for severe cases [3,5]. However, it might be difficult by only through nutritional interventions to improve or prevent sarcopenia. The benefit of physical activity and exercise is well known for sarcopenia, and the international clinical practice guidelines for sarcopenia in 2018 strongly recommend physical activity for the treatment of sarcopenia [27]. Any single intervention in the aspect of nutrition, exercise, or medication is less effective for the management of sarcopenia, and proper dietary intake and exercise should be combined [28]. Resistance exercise has been shown to improve muscle mass and strength [29,30].

An interesting result was the number of weekly walk steps. The mean of the walk steps reached a maximum of 20,000 in the second week and a minimum of 7000 in the 6th week. It is assumed that the limited physical activity during the fourth to the 6th week could cause muscle loss. This pattern is associated with the severity of radiation esophagitis. In general, RT esophagitis begins at week 2 of RT, its severity increases over time, and peaks at the end of RT. It takes approximately 3 weeks for the symptoms to resolve. Detailed analyses of the walking steps showed that patients with excessive muscle loss during NACRT significantly decreased the number of steps in the last half of the trial period. Furthermore, age affected the number of walking steps, and pre-RT sarcopenia showed a difference in the recovery of the decreased walking steps after NACRT. This means that older patients and those with sarcopenia at the start of the treatment have difficulty maintaining physical activity during the treatment and recovering decreased physical activity after NACRT. Some studies suggested that the number of walking steps was related to performance status [31]. As performance status is also one of the important factors for surgical indication and outcome [32], the pattern of the decrease and recovery in walking steps after NACRT is likely to be an indicator for the continuation of follow-up surgery or a predictor for postoperative mortality in patients with esophageal cancer.

To prevent malnutrition and muscle loss in esophageal cancer patients receiving NACRT, comprehensive supportive care including nutritional care as well as exercise might be needed. In addition, the status of nutrition and muscle mass needs to be evaluated individually before the start of treatment. For example, patients without malnutrition and sarcopenia can be managed with the mobile interactive coaching app used in this study. However, patients with old age, malnutrition, and sarcopenia might need to be managed differently. Physical activity guidelines for cancer patients from the American Cancer Society recommend avoiding inactivity, continuing normal daily activity as much as possible, and performing exercise adapted for the disease and the patients' condition, particularly if patients experience severe symptoms such as extreme fatigue or ataxia, or have cardiovascular and pulmonary contraindications [33]. However, exercise programs targeting patients with sarcopenia or extremely low physical activity are not well established, and high-intensity exercise might be not good for patients undergoing cancer treatment. It is necessary to develop a customized and intensity-modified physical program for patients with malnutrition and sarcopenia.

As a limitation, this prospective study was a small pilot test, thus limiting detailed analyses. We used the SMI from CT to diagnose sarcopenia, but muscle strength, which is one of the criteria for the definition of sarcopenia, could not be evaluated. In addition, all enrolled patients were male for comparison with our previous study, which predominantly involved esophageal cancer patients for ensuring study homogeneity. Finally, we focused on the app usage such as the activation or automatically recorded walk steps, but more specific nutritional factors such as calories during oral intake were missing. Although the app was not highly effective in preventing muscle loss and further research using more detailed information is needed, we expect that it can be used as a self-managing nutritional support for cancer patients and possibly expanded for cancer survivors after treatment.

Conclusion

For esophageal cancer patients receiving NACRT, an interactive health coaching mobile app helped nutritional self-care with a significantly less decrease in PNI, although it did not prevent excessive muscle loss. Low physical activity estimated by the number of walking steps did not recover even a few weeks after the end of NACRT in patients with old age or pretreatment sarcopenia. An individualized care model with proper exercise as well as nutritional support may be required to reduce muscle loss and malnutrition.

Acknowledgments

Noom Inc provided full range of services in the Noom app during this trial.

Conflicts of Interest

Although HK, JL, and YK are affiliated to the company providing the app, this study was based on only objective measures such as patient activities using the app, walk steps, and clinical data. The other authors have no conflict of interest to disclose.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018 Nov;68(6):394-424 [FREE Full text] [doi: 10.3322/caac.21492] [Medline: 30207593]
2. van Hagen P, Hulshof M, van Lanschot J, Steyerberg E, Henegouwen MVB, Wijnhoven B, et al. Preoperative chemoradiotherapy for esophageal or junctional cancer. N Engl J Med 2012 May;366(22):2074-2084. [doi: 10.1056/nejmoa1112088]
3. Jordan T, Mastnak DM, Palamar N, Kozjek NR. Nutritional therapy for patients with esophageal cancer. Nutr Cancer 2017 Oct;70(1):23-29. [doi: 10.1080/01635581.2017.1374417] [Medline: 29016197]
4. Kirby DF, Teran JC. Enteral feeding in critical care, gastrointestinal diseases, and cancer. Gastrointest Endosc Clin N Am 1998 Jul;8(3):623-643. [Medline: 9654573]
5. Meier R, Forbes A. Basics in clinical medical nutrition. In: The Importance of Nutrition as an Integral Part of Disease Management. Basel: Karger; 2015 Presented at: 82nd Nestlé Nutrition Institute Workshop; October 2015; New Delhi p. 1-16. [doi: 10.1159/000381997]
6. Anandavadivelan P, Lagergren P. Cachexia in patients with oesophageal cancer. Nat Rev Clin Oncol 2015 Nov;13(3):185-198. [doi: 10.1038/nrclinonc.2015.200]
7. Yoon HG, Oh D, Ahn YC, Noh JM, Pyo H, Cho WK, et al. Prognostic impact of sarcopenia and skeletal muscle loss during neoadjuvant chemoradiotherapy in esophageal cancer. Cancers (Basel) 2020 Apr;12(4):925 [FREE Full text] [doi: 10.3390/cancers12040925] [Medline: 32290037]
8. https://www.noom.com. URL: https://www.noom.com [accessed 2021-08-04]
9. Chin SO, Keum C, Woo J, Park J, Choi HJ, Woo J, et al. Successful weight reduction and maintenance by using a smartphone application in those with overweight and obesity. Sci Rep 2016 Nov;6:34563 [FREE Full text] [doi: 10.1038/srep34563] [Medline: 27819345]
10. Jacobs S, Radnitz C, Hildebrandt T. Adherence as a predictor of weight loss in a commonly used smartphone application. Obes Res Clin Pract 2017 Mar;11(2):206-214. [doi: 10.1016/j.orcp.2016.05.001] [Medline: 27292942]
11. Gallagher D, Rosenn B, Toro-Ramos T, Paley C, Gidwani S, Horowitz M, et al. Greater neonatal fat-free mass and similar fat mass following a randomized trial to control excess gestational weight gain. Obesity 2018 Feb;26(3):578-587. [doi: 10.1002/oby.22079]
12. Michaelides A, Major J, Pienkosz E, Wood M, Kim Y, Toro-Ramos T. Usefulness of a novel mobile diabetes prevention program delivery platform with human coaching: 65-week observational follow-up. JMIR Mhealth Uhealth 2018 May;6(5):e93 [FREE Full text] [doi: 10.2196/mhealth.9161] [Medline: 29724709]

13. Hildebrandt T, Michaelides A, Mackinnon D, Greif R, DeBar L, Sysko R. Randomized controlled trial comparing smartphone assisted versus traditional guided self-help for adults with binge eating. Int J Eat Disord 2017 Nov;50(11):1313-1322 [FREE Full text] [doi: 10.1002/eat.22781] [Medline: 28960384]
14. Toro-Ramos T, Michaelides A, Anton M, Karim Z, Kang-Oh L, Argyrou C, et al. Mobile delivery of the diabetes prevention program in people wth prediabetes: randomized controlled trial. JMIR Mhealth Uhealth 2020 Jul;8(7):e17842 [FREE Full text] [doi: 10.2196/17842] [Medline: 32459631]
15. https://sourceforge.net/projects/muscle-fatarea-measurement. URL: https://sourceforge.net/projects/muscle-fatarea-measurement [accessed 2021-08-04]
16. Prado CMM, Lieffers JR, McCargar LJ, Reiman T, Sawyer MB, Martin L, et al. Prevalence and clinical implications of sarcopenic obesity in patients with solid tumours of the respiratory and gastrointestinal tracts: a population-based study. Lancet Oncol 2008 Jul;9(7):629-635. [doi: 10.1016/S1470-2045(08)70153-0] [Medline: 18539529]
17. Migita K, Takayama T, Saeki K, Matsumoto S, Wakatsuki K, Enomoto K, et al. The prognostic nutritional index predicts long-term outcomes of gastric cancer patients independent of tumor stage. Ann Surg Oncol 2013 Mar;20(8):2647-2654. [doi: 10.1245/s10434-013-2926-5]
18. http://www.R-project.org. URL: http://www.R-project.org [accessed 2021-08-04]
19. Santilli V, Bernetti A, Mangone M, Paoloni M. Clinical definition of sarcopenia. Clin Cases Miner Bone Metab 2014 Sep;11(3):177-180 [FREE Full text] [Medline: 25568649]
20. Rosenberg IH. Sarcopenia: origins and clinical relevance. Clin Geriatr Med 2011 Aug;27(3):337-339. [doi: 10.1016/j.cger.2011.03.003] [Medline: 21824550]
21. Gale CR, Martyn CN, Cooper C, Sayer AA. Grip strength, body composition, and mortality. Int J Epidemiol 2007 Mar;36(1):228-235. [doi: 10.1093/ije/dyl224] [Medline: 17056604]
22. Chindapasirt J. Sarcopenia in cancer patients. Asian Pac J Cancer Prev 2015 Jan;16(18):8075-8077 [FREE Full text] [doi: 10.7314/apjcp.2015.16.18.8075] [Medline: 26745041]
23. Roy P, Chowdhury S, Roy HK. Exercise-induced myokines as emerging therapeutic agents in colorectal cancer prevention and treatment. Future Oncol 2018 Feb;14(4):309-312 [FREE Full text] [doi: 10.2217/fon-2017-0555] [Medline: 29318900]
24. Lin J, Zhang W, Huang Y, Chen W, Wu R, Chen X, et al. Sarcopenia is associated with the neutrophil/lymphocyte and platelet/lymphocyte ratios in operable gastric cancer patients: a prospective study. CMAR 2018 Oct;10:4935-4944. [doi: 10.2147/cmar.s175421]
25. Foucaut A, Morelle M, Kempf-Lépine AS, Baudinet C, Meyrand R, Guillemaut S, et al. Feasibility of an exercise and nutritional intervention for weight management during adjuvant treatment for localized breast cancer: the PASAPAS randomized controlled trial. Support Care Cancer 2019 Sep;27(9):3449-3461. [doi: 10.1007/s00520-019-4658-y] [Medline: 30680617]
26. Rinninella E, Cintoni M, Raoul P, Pozzo C, Strippoli A, Bria E, et al. Effects of nutritional interventions on nutritional status in patients with gastric cancer: a systematic review and meta-analysis of randomized controlled trials. Clin Nutr ESPEN 2020 Aug;38:28-42. [doi: 10.1016/j.clnesp.2020.05.007] [Medline: 32690170]
27. Dent E, Morley JE, Cruz-Jentoft AJ, Arai H, Kritchevsky SB, Guralnik J, et al. International Clinical Practice Guidelines for Sarcopenia (ICFSR): screening, diagnosis and management. J Nutr Health Aging 2018 Nov;22(10):1148-1161. [doi: 10.1007/s12603-018-1139-9] [Medline: 30498820]
28. Cruz-Jentoft AJ, Sayer AA. Sarcopenia. The Lancet 2019 Jun;393(10191):2636-2646. [doi: 10.1016/s0140-6736(19)31138-9]
29. Peterson MD, Rhea MR, Sen A, Gordon PM. Resistance exercise for muscular strength in older adults: a meta-analysis. Ageing Res Rev 2010 Jul;9(3):226-237 [FREE Full text] [doi: 10.1016/j.arr.2010.03.004] [Medline: 20385254]
30. Peterson M, Sen A, Gordon P. Influence of resistance exercise on lean body mass in aging adults: a meta-analysis. Med Sci Sports Exerc 2011 Feb;43(2):249-258 [FREE Full text] [doi: 10.1249/MSS.0b013e3181eb6265] [Medline: 20543750]
31. Gresham G, Hendifar AE, Spiegel B, Neeman E, Tuli R, Rimel BJ, et al. Wearable activity monitors to assess performance status and predict clinical outcomes in advanced cancer patients. NPJ Digit Med 2018;1:27 [FREE Full text] [doi: 10.1038/s41746-018-0032-6] [Medline: 31304309]
32. Cihoric M, Tengberg LT, Foss NB, Gögenur I, Tolstrup M, Bay-Nielsen M. Functional performance and 30-day postoperative mortality after emergency laparotomy-a retrospective, multicenter, observational cohort study of 1084 patients. Perioper Med (Lond) 2020 May;9(1):13 [FREE Full text] [doi: 10.1186/s13741-020-00143-7] [Medline: 32391145]
33. Schmitz KH, Courneya KS, Matthews C, Demark-Wahnefried W, Galvão DA, Pinto BM, et al. American College of Sports Medicine roundtable on exercise guidelines for cancer survivors. Med Sci Sports Exerc 2010 Jul;42(7):1409-1426. [doi: 10.1249/MSS.0b013e3181e0c112] [Medline: 20559064]

Abbreviations

ALC: absolute lymphocyte count
ANC: absolute neutrophil count
ESOG PS: Eastern Cooperative Oncology Group Performance Status

ESPEN-SIG: European Society for Clinical Nutrition and Metabolism Special Interest Groups IWGS: International Working Group on Sarcopenia
EWGSOP: European Working Group on Sarcopenia in Older People
NACRT: neoadjuvant chemoradiotherapy
NLR: neutrophil-to-lymphocyte ratio
PNI: prognostic nutritional index
RT: radiotherapy
SMI: skeletal muscle index
WBC: white blood cell

Edited by R Kukafka; submitted 11.03.21; peer-reviewed by SG Park, M Hamidzadeh; comments to author 19.04.21; revised version received 30.04.21; accepted 13.07.21; published 27.08.21.

Please cite as:
Yang K, Oh D, Noh JM, Yoon HG, Sun JM, Kim HK, Zo JI, Shim YM, Ko H, Lee J, Kim Y
Feasibility of an Interactive Health Coaching Mobile App to Prevent Malnutrition and Muscle Loss in Esophageal Cancer Patients Receiving Neoadjuvant Concurrent Chemoradiotherapy: Prospective Pilot Study
J Med Internet Res 2021;23(8):e28695
URL: https://www.jmir.org/2021/8/e28695
doi:10.2196/28695
PMID:34448714

©Kyungmi Yang, Dongryul Oh, Jae Myoung Noh, Han Gyul Yoon, Jong-Mu Sun, Hong Kwan Kim, Jae Ill Zo, Young Mog Shim, Hyunyoung Ko, Jungeun Lee, Youngin Kim. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Perceptions of Alerts Issued by Social Media Platforms in Response to Self-injury Posts Among Latinx Adolescents: Qualitative Analysis

Linnea I Laestadius[1], MPP, PhD; Katherine A Craig[2], MA; Celeste Campos-Castillo[3], PhD

[1]Zilber School of Public Health, University of Wisconsin-Milwaukee, Milwaukee, WI, United States
[2]Department of Sociology, University of Colorado Boulder, Boulder, CO, United States
[3]Department of Sociology, University of Wisconsin-Milwaukee, Milwaukee, WI, United States

Corresponding Author:
Linnea I Laestadius, MPP, PhD
Zilber School of Public Health
University of Wisconsin-Milwaukee
P O Box 413
Milwaukee, WI, 53201
United States
Phone: 1 414 251 5607
Email: llaestad@uwm.edu

Abstract

Background: There is growing interest in using social media data to detect and address nonsuicidal self-injury (NSSI) among adolescents. Adolescents often do not seek clinical help for NSSI and may adopt strategies to obscure detection; therefore, social media platforms may be able to facilitate early detection and treatment by using machine learning models to screen posts for harmful content and subsequently alert adults. However, such efforts have raised privacy and ethical concerns among health researchers. Little is currently known about how adolescents perceive these efforts.

Objective: The aim of this study is to examine perceptions of automated alerts for NSSI posts on social media among Latinx adolescents, who are at risk for NSSI yet are underrepresented in both NSSI and health informatics research. In addition, we considered their perspectives on preferred recipients of automated alerts.

Methods: We conducted semistructured, qualitative interviews with 42 Latinx adolescents between the ages of 13 and 17 years who were recruited from a nonprofit organization serving the Latinx community in Milwaukee, Wisconsin. The Latinx population in Milwaukee is largely of Mexican descent. All interviews were conducted between June and July 2019. Transcripts were analyzed using framework analysis to discern their perceptions of automated alerts sent by social media platforms and potential alert recipients.

Results: Participants felt that automated alerts would make adolescents safer and expedite aid before the situation escalated. However, some worried that hyperbolic statements would generate false alerts and instigate conflicts. Interviews revealed strong opinions about ideal alert recipients. Parents were most commonly endorsed, but support was conditional on perceptions that the parent would respond appropriately. Emergency services were judged as safer but inappropriate for situations considered lower risk. Alerts sent to school staff generated the strongest privacy concerns. Altogether, the preferred alert recipients varied by individual adolescents and perceived risks in the situation. None raised ethical concerns about the collection, analysis, or storage of personal information regarding their mental health status.

Conclusions: Overall, Latinx adolescents expressed broad support for automated alerts for NSSI on social media, which indicates opportunities to address NSSI. However, these efforts should be co-constructed with adolescents to ensure that preferences and needs are met, as well as embedded within broader approaches for addressing structural and cultural barriers to care.

(J Med Internet Res 2021;23(8):e28931) doi:10.2196/28931

KEYWORDS

adolescents; social media; mental health; NSSI; race and ethnicity; mobile phone

Introduction

Background

The rates of nonsuicidal self-injury (NSSI) and suicidal ideation have been rising among adolescents in the United States, with Latinx adolescents experiencing high rates of both [1-3]. NSSI, also known as self-harm or self-injury, is defined as “deliberate, self-inflicted damage of body tissue without suicidal intent and for purposes not socially or culturally sanctioned” [4]. Although distinct in intent, NSSI is a risk factor for suicidality [5,6]. There is evidence that social media users share web-based NSSI content [7,8], creating interest in approaches using social media data to detect and address NSSI and prevent suicidality [9-13].

One approach, termed digital phenotyping, suggests that analysis (usually via machine learning) of data generated from interactions with digital technologies and devices, including social media, can be used to “identify and diagnose health conditions” [14-16]. Although initially envisioned to include clinician-generated inputs, such as medical records, the approach is now regularly applied to social media data without further inputs [10,16-19]. As adolescents often do not seek clinical help for NSSI [20] and may adopt strategies to obscure detection [7,21-23], the creation of alerts informed by digital phenotyping of social media data could provide new opportunities for early detection and treatment.

Although a potentially promising intervention, digital phenotyping alerts based on machine learning algorithms to detect mental health concerns raise questions about consent, efficacy, and privacy risks [14,24]. For example, Facebook uses digital phenotyping to detect suicidal ideation in posts made by US users and then alerts local emergency responders when the staff deems it is appropriate [13]. There is no way to opt into or out of this feature, even for minors [25]. Some have also raised questions about the accuracy of the underlying machine learning model analyzing posts to infer mental health status, particularly in cases where predictions are primarily built using data from adults [26]. Both false negatives and false positives weaken the efficacy of the intervention and may pose risks to adolescents. Although adults have shown little support for these efforts [27], adolescents’ perceptions of both alerts based on their social media posts and the commercial and institutional uses of personal data remain largely unknown [28]. Given the initial evidence that youth are less likely than adults to view their health information as sensitive and the collection of their personal data by companies as a privacy violation [29], they may be apathetic toward these interventions. However, the only prior study of the perceptions of young adults of digital phenotyping (which focused on assessing mental health via cellphone sensor data) demonstrated that they had concerns about privacy and autonomy and rejected the monitoring of text message content [30].

As interventions informed by social media data are already deployed and will likely spread to an increasing number of settings, it is an ethical and design imperative to understand how adolescents perceive these efforts and how mental health alerts based on digital phenotyping can best be tailored to meet their preferences and needs [12]. Prior research has stressed the importance of user-centered design and the cocreation of digital mental health tools with youth [25,30]. Moving ahead without considering youth perspectives may result in both rights violations and backlash against existing and future interventions [25]. For example, the only way adolescents may opt out of Facebook’s suicide intervention efforts is to discontinue using the Facebook and Instagram platforms, which may have negative implications for adolescents and would limit the completeness of data used to inform subsequent interventions. Accordingly, assessing youth perspectives on how interventions are designed may help improve their efficacy and accuracy.

Study Aims

We address this gap in the literature by qualitatively examining how Latinx adolescents perceive social media platforms automatically detecting posts about NSSI and alerting an authority figure in response (*automated alerts*). We focus on Latinx adolescents because although they engage in NSSI, they are underrepresented in research on NSSI and are more likely to be undiagnosed than White adolescents [31]. Latinx adolescents also experience higher rates of depression, which is a risk factor for NSSI [32,33], than non-Hispanic White adolescents [2,32-34]. The actual rates of depression may be higher because clinicians may miss cultural differences in the expression of symptoms [35]. Thus, population-level models may introduce cultural bias into digital phenotyping efforts to infer NSSI [36]. At the same time, Latinx adolescents may stand to benefit from such tools because of socioeconomic and cultural barriers to receiving mental health screening [37]. Moreover, the Latinx community is underrepresented in studies of privacy on social media [38-40] despite factors such as documentation status and larger household sizes, potentially creating unique privacy needs and ethical concerns [41]. In the context of mental health, potential stigmatization shapes who they are likely to approach for mental health support [42] and thus may also influence their preferences for who should receive an alert. Accordingly, it is critical to engage with Latinx adolescents to study how they perceive interventions that automatically alert a third party to obtain help when adolescents make concerning social media posts and their preferences for alert recipients.

Methods

Data Collection

Eligible participants were Latinx adolescents aged 13 to 17 years who used social media, defined as posting at least once per week. We collaborated with a community organization serving the Latinx community in Milwaukee, Wisconsin, which hosts a summer youth program to recruit and screen participants based on the eligibility criteria. Adolescents gave assent, and guardians provided consent before data collection proceeded. Both assent and consent documents were available in English and Spanish. Although the family origin of students at the community organization was not available, Milwaukee’s Latinx population is largely of Mexican descent [43]. Participants who were enrolled in the study received a US $40 incentive. The institutional review board at the University of Wisconsin-Milwaukee approved the research protocol.

Between June and July 2019, 60- to 90-minute semistructured, in-person interviews were held with 43 Latinx adolescents in a quiet room at the community organization. Although a Spanish-speaking interviewer was available, all interviews were conducted in English at the request of the adolescents. At the start of the interview, participants provided demographic information and selected a study pseudonym. As part of a broader line of questioning about how adolescents handle mental health disclosures from peers on social media [44], participants were asked the following question: "How would you feel if a social media app sent an automatic alert to an adult when you or your friends made a post about self-harm?" The term self-harm was chosen over NSSI to reflect common use [7]. The participants were then asked who should receive such alerts. We probed about emergency services, which the Facebook suicide intervention alerts, and *trusted adults* (parents or guardians and school staff or teachers) frequently identified in prior literature [44,45]. Participants were also probed about why they endorsed or opposed these alert recipients. One interview concluded without asking about automated alerts because of time constraints, resulting in a sample size of 42 participants.

Data Analysis

Following professional verbatim transcription, the interdisciplinary team, consisting of a senior qualitative researcher in public health, a senior sociologist who is Latinx, and 2 graduate students trained in qualitative methods, analyzed the transcripts using framework analysis [46] to identify perceptions of automated NSSI alerts and alert recipients. The 2 senior members performed open coding on the transcripts for the first 12 interviews to identify the initial codes and develop the codebook. The codebook captured perceptions for and against automated alerts. Transcripts were then split among team members and coded using MAXQDA 2018 (VERBI Software GmbH). Throughout the final coding process, an additional 11 transcripts were coded by 2 to 3 members and discussed code by code to allow new codes to arise and ensure a shared understanding of code definitions. Any discrepancies were resolved through consensus, with prior coding updated as needed using an iterative process [47]. Coded transcripts were then charted into a framework analysis matrix [46] using MAXQDA 2018 to summarize data by category from each transcript, with a focus on identifying favorable and unfavorable perceptions of automated alerts and alert recipients. In addition, memos were written to aid in the data interpretation and identification of representative quotes [48].

Results

Overview

Of the 42 participants, 24 (57%) self-identified as girls and 18 (43%) as boys. The ages ranged from 13 to 17 years, with a mean age of 15. Over 61% (26/42) of the participants indicated that both of their parents were immigrants in the United States. The participants most frequently used Snapchat (42/42, 100%), followed by Instagram (39/42, 93%). We present findings on the perceptions of automated alerts for NSSI, followed by the preferred recipient of such alerts. The majority of participants felt that automated alerts would make adolescents safer, but also expressed strong concerns about alert accuracy and the efficacy of alert recipients. The representative quotes below are presented using the pseudonyms selected by the participants.

How do Latinx Adolescents Perceive Automated Alerts for NSSI?

Participants worried about whether digital phenotyping tools could accurately detect NSSI, specifically mentioning false positives due to (1) the frequent use of jokes, sarcasm, and hyperbole by adolescents, and (2) situations that were seen as posing only minor levels of risk that did not necessitate intervention. For example, Dahlia (age 14) supported alerts but worried about social media posts being misinterpreted, and alerts sent inappropriately:

> *I think I would be okay with that because in the end they're just trying to help, but at the same time if like you were saying that but you didn't really mean that and someone were to find out, that would be like embarrassing.*

Ariona (age 14) similarly worried about common turns of speech being taken literally and parents overreacting to an alert:

...if it was just joking around, your parents might take it way too seriously. [Ariona]

So what's an example of someone joking around about self-harm? [Interviewer]

Like sometimes when we have a bad test that was really hard and we'll be like, "I want to kill myself," but it's just joking. We're not being for real about it. [Ariona]

Fears also included getting in trouble with authorities if emergency services were summoned without an emergency being present.

Pablo (age 17) explained as follows:

> *Don't you get like fined and stuff like that like if the ambulance comes?...What if the algorithm gets it wrong for some reason then you get some ambulances outside your house?*

Broad privacy concerns were relatively rare and focused primarily on the desire for adolescents to tell adults themselves rather than an impersonal alert being sent on their behalf. Although none of the participants indicated that they would stop posting to regain control of their privacy, a small number anticipated that other adolescents may do so. More commonly, participants dismissed privacy concerns as minor relative to the severe risks posed by NSSI. Although the interview questions about automated alerts focused on self-harm, participants often linked NSSI with suicidality, both conceptually (as with the quote from Ariona above) and with regard to the importance of timely intervention to curtail escalation. Perceptions of urgency were the primary drivers of support for automated alerts. Ashley (age 15) explained as follows:

> *Like it's good for [parents] to get that alert just so you don't do anything more stupid than trying to self-harm yourself, and so that way your parents can know on time to like hurry and like tell you before you do anything else.*

Similarly, David (age 17) felt alerts would “probably be a good thing because it’d probably alert the parent or guardian before the harmful thought got too serious.”

Participants also noted that adolescents often post about mental health struggles without anyone acting on it and expressed a strong sentiment that parents or guardians are commonly unaware of how their children are feeling. Bri (age 17) felt that alerts were as follows:

> *pretty smart, because so many people tend to stay quiet. They tend to keep it on social media. They usually don’t tell their parents, hey, this person that I know is feeling like this. And it’s just better to be safe than sorry...It’s better to do that than say that and then no one says anything, and then it just stays in that toxic state. Like it’s just there. You keep saying you’re depressed, and no one says anything. And something could happen.*

Consequently, participants felt that adolescents struggling with NSSI may experience barriers to timely help. Automated alerts were seen as faster and more reliable for obtaining aid than waiting on friends to take action or parents or guardians to notice. Alerts were mentioned as a means of improving communication among adults and children. Once notified, participants generally anticipated that adults would offer adolescents some form of mental health support. Monse Marie (age 13) explained this as follows:

> *I think that would be a good thing for the social media to do because if you’re going through something then the parent will try to help you like oh, maybe therapy or something, you know, besides like because sometimes parents are blinded, like they don’t know what’s going on with the kid.*

Who Should Receive Automated Alerts for NSSI?

Despite strong support for automated alerts because they could yield adult help, participants were cognizant of adults differing in their efficacy and desire to help. An alert sent to the wrong person could fail to generate help and instead cause negative repercussions if sent to someone who would punish rather than help the adolescent. Although most participants suggested that alerts should be sent to parents or guardians because of their close bonds to their children, parents or guardians also stood out in interviews as both unpredictable and occasionally hurtful. Lucy (age 14) recalled the example of a friend who illustrated this as follows:

> *...my friend, she’s felt sad before and she even mentioned depression once, and her parents got really mad at her, and they blamed it on her that she was feeling sad and that she mentioned it. So I think that it would be best to tell a teacher versus a parent, or tell emergency services...*

John (age 14) similarly noted that “if it were an abusive parent or something like that it wouldn’t be good.” A small number of participants also expressed fear of how their own parents would respond to the alerts.

One common concern was that parents or guardians would not take mental health concerns seriously and may not know how to respond. Lilo (age 17) preferred emergency services over parents since:

> *self-harm is something that somebody should get treated for and their parent might...be passive about it. Like you never know how serious a parent is going to take it.*

Bri (age 17) felt that Latinx families specifically may have limited efficacy in handling mental health concerns:

> *I feel like it’s different, because of the fact that I know, for Latinos, when there’s mental health included, it’s really hard to talk about it with your parents, especially if your parent is not educated in that, or if your parents is—they’ve grown up in certain ways, and so they will believe that what we think is kind of dumb. Because I know, for some families, they can’t even talk about that type of thing, because it’s seen as a weakness.*

Emergency services were more uniformly seen as safe and knowledgeable, but participants felt that not all posts indicating NSSI merited intervention from emergency services. Some participants mentioned a hotline number as an alternative. No participant worried about emergency services introducing risks to their families, although one stated “you wouldn’t want the FBI knocking on your door” [John, age 14].

School staff and teachers were occasionally mentioned as preferable substitutions when participants thought parents or guardians would be unsupportive, but these adults were not immune from concerns about unhelpful responses and punishment. Although privacy concerns for automated alerts in general were minimal, privacy concerns specific to alerts sent to schools were more common. Specifically, some participants worried about other students finding out or that school staff and teachers would judge adolescents for their mental health struggles. Jaden (age 16) explained as follows:

> *I wouldn’t want that business out there to my school if, you know, right away. Because if they don’t understand the context of the situation, that could just also do the whole labeling thing where it’s like you feel like everyone’s kind of just watching out for you...I wouldn’t want that business there with my school because it’s just, it’s a big place and you’re there for a long time, so it’s kind of like it could have lasting effects on you while you’re there.*

Although preferred alert recipients differed across individual adolescents and across perceived severity of situations, participants shared a desire that alerts should be sent to “someone who cares and who wants them to get better” [Sara, age 14]. Overall, the alert recipient was viewed as critical to the efficacy and desirability of alerts derived from social media posts.

Discussion

Principal Findings

This study examined the perceptions of Latinx adolescents toward automated alerts for NSSI based on digital phenotyping using social media platforms. Diverging from adult perceptions [27], participants were largely positive toward automated alerts, suggesting that alerts could help overcome the lack of responsiveness from peers and parents. In particular, participants felt that alerts could connect adolescents to help before situations escalated. Notably, the perception that social media users often ignore posts with concerning content rather than report it is among the reasons that Facebook launched its digital phenotyping efforts [13]. Adolescents stressed that alerts must be accurate and directed toward someone able and willing to help. The efficacy of alerts as a mental health intervention is partially contingent on the efficacy of the alert recipient, suggesting the need to adopt practices that accommodate and support the social fabric surrounding adolescents.

The findings reveal opportunities to improve the ethics and efficacy of mental health interventions based on digital phenotyping. With respect to the machine learning algorithms that underpin alerts, the findings show that these must be sensitive to the specific linguistic patterns of adolescents to account for the use of sarcasm, jokes, and hyperbole used for dramatic effect. Furthermore, algorithms must be sensitive to variation across adolescents, including those rooted in cultural differences in the expression of depression and NSSI [35]. Although participants did not suggest that they had concerns unique to their status as Latinx youth, attention should be paid to the specific subpopulations included in the data to train algorithms [36]. Although continued manual screening of posts identified through machine learning, similar to the efforts of Facebook, may help reduce the number of alerts being sent inappropriately, poorly trained machine learning tools are still likely to miss concerning posts. In line with prior research showing that depression can co-occur with NSSI [32,33], interviewees relayed how detecting depression early could offer timely help to limit adolescent harm.

With regard to the actual alerts, the selection of an appropriate recipient is crucial to the adolescent receiving assistance. NSSI represents a particularly interesting scenario for alerts because, unlike suicidality, emergency services are not always the obvious or appropriate choice. Although many participants either conflated NSSI with suicidality or worried about NSSI leading to something more serious, which accords with prior research on Latinx adolescents [31], they felt that emergency services may be an excessive response for NSSI. However, not all adolescents reported access to supportive school staff, teachers, parents, or guardians, and some worried adults would respond in detrimental ways. These concerns are consistent with those found in prior research on adolescents who engage in NSSI and highlight the importance of increased education for adults who surround adolescents [49].

As there was no uniformly appropriate alert recipient across all participants and levels of need, researchers and social media platform developers should consider the development of alerts that take into account the privacy preferences of *individual* adolescents. The lack of uniformly appropriate alert recipients is consistent with a contextual view of privacy [50] and suggests that interventions should be co-designed with adolescents providing feedback to ensure that efforts meet their preferences and needs [51]. In the implementation stage, adolescents should be asked whether they want alerts to be sent for NSSI and, if so, to whom. Engagement is ideally an ongoing process, as privacy preferences and trusted adults may change over time. Allowing participants to opt in to alerts would address one of the major ethical concerns with the current approach of Facebook to digital phenotyping [25].

At the time of writing, we were not aware of any platforms that allow social media users to choose in advance who should receive alerts sent by platforms. Facebook makes use of a two-tiered system in which their machine learning tool identifies users who may be at risk of imminent harm. In the event that Facebook's Community Operations team agrees with the assessment, the user is shown *support options* [52], including a prompt to reach out to a Facebook friend using prepopulated text. However, the user must decide in that moment to send the message themselves. In the event that the risk to the user is deemed high, the Facebook team instead notifies local emergency services. It should be noted, however, that this system was designed for suicidality rather than NSSI. Furthermore, other platforms popular with adolescents do not have an official alert policy at all, even for suicidality. TikTok, for example, simply states that those considering harming themselves should "please contact [their] local emergency services, a suicide-prevention helpline, or expert listening service for help" [53]. Platforms have room for significant growth in the area of co-designing responses for both NSSI and suicidality. It will also be critical to engage parents, teachers, and emergency service managers in this co-design process to examine their preferences for alerts. In addition, further research is needed to examine and support the capacity of these actors to manage alerts in a constructive manner, especially with respect to recognizing when and how to seek mental health treatment. As noted by the participants themselves, the notification of an adult is not a guarantee that an adolescent will receive mental health services.

Although participants mentioned concerns about their privacy within school settings, none voiced concerns related to social media platforms analyzing or storing mental health relevant information, or ethical concerns about this occurring without their awareness or explicit consent. Furthermore, none expressed concerns about the implications of these data for future employment or insurability [24]. Although there is little research on adolescent perceptions of commercial and institutional privacy [28], these findings are consistent with initial evidence that adolescents are less likely than adults to view their health information as sensitive [29]. It is also possible that the absence of concerns reflects a lack of awareness of risks. Adolescents in the United States may be at particular risk for harms stemming from the intentional or inadvertent disclosure of private health-related information by platforms in light of the power of the private health insurance industry. Although the Affordable Care Act of 2010 currently prevents insurers from denying coverage or charging higher rates due to preexisting conditions, these provisions are currently in jeopardy because of legal challenges [54]. Future research should provide participants with risk information and assess whether support changes for social media platforms engaging in digital phenotyping. Any effort by social media companies to expand current efforts should adopt a digital literacy component to achieve meaningful informed consent. More broadly, it is notable that Facebook's automated suicide-prevention feature is not available in the European Union because the General Data Protection Regulation provides social media users greater control over their *data concerning health* [12,55]. The United States currently has no equivalent law, and the Health Insurance Portability and Accountability Act applies only to health care providers, plans, and clearinghouses [55]. Further work is needed to ensure that social media platforms protect the health-related information they gather and analyze, particularly concerning minors.

Overall, the findings suggest that automated alerts based on digital phenotyping of social media data may offer opportunities to address NSSI among Latinx adolescents and prevent further harm. This is a significant finding given the high rates of NSSI and barriers to receiving mental health treatment among Latinx adolescents [2,35,56]. Further, Latinx adolescents have a high rate of smartphone access and report high levels of social media use [57], indicating a wide range of opportunities for deployment. However, the implementation of alerts without addressing underlying disparities in access to mental health services in the United States may exacerbate disparities because such interventions risk being more effective for populations with access to care [37,58-60]. In addition to practical concerns with regard to algorithms and alert recipients, developers should also be mindful of the broader socioeconomic implications of their efforts.

Limitations

With regard to limitations, the transferability of findings should be considered in light of the fact that this study was conducted with English-speaking Latinx adolescents drawn from a summer program in a midsized US city that is largely of Mexican origin [43]. The homogeneous sample helps address the underrepresentation of Latinx groups in privacy research and their disproportionate risk of NSSI. Further research is needed to reveal how concerns and preferences may be distinct across populations. In addition, as this study was one element within a larger project, time constraints prevented additional questions prompting participants to reflect explicitly on questions of commercial privacy. However, the absence of probing allowed us to capture the degree to which participants readily reflected on these concerns. It should also be stressed that participants were not selected based on prior experiences with NSSI, which allows findings to be transferable to a broader pool of adolescents but may fail to capture the unique preferences of those engaging in NSSI. In light of prior research indicating that many adolescents choose to seek help for NSSI on the internet [61], and that these adolescents may experience more suicidality than those who do not seek help on the internet [62], there is a strong need for additional work exploring whether the benefits of NSSI alerts would offset any potential chilling effects on help seeking via social media. Finally, future research should also expand beyond adolescents to consider the capacity and preferences of adults and institutions who could serve as alert recipients.

Conclusions

This research finds broad support among Latinx adolescents for automated alerts issued in response to NSSI posts on social media, indicating an opportunity for deploying interventions grounded in digital phenotyping to address mental health on the internet. Latinx adolescents described pragmatic concerns about the accuracy of these tools and who should receive alerts but had few ethical concerns and expressed little awareness of the risks posed by the collection and storage of health-related information by commercial entities. Researchers and social media platforms should collaborate with adolescents to co-design interventions that accurately and appropriately alert a chosen recipient when they are at risk for NSSI. Any such effort should consider the implications for adolescent privacy. Ensuring access to appropriate mental health services following detection will be critical to avoid exacerbating disparities.

Acknowledgments

This study was funded by a grant from the Technology and Adolescent Mental Wellness program at the University of Wisconsin-Madison. The content is solely the responsibility of the authors and does not necessarily represent the official views of the university or the Technology and Adolescent Mental Wellness program.

Conflicts of Interest

None declared.

References

1. Zwald ML, Holland KM, Annor FB, Kite-Powell A, Sumner SA, Bowen DA, et al. Syndromic surveillance of suicidal ideation and self-directed violence - United States, January 2017-December 2018. MMWR Morb Mortal Wkly Rep 2020 Jan 31;69(4):103-108 [FREE Full text] [doi: 10.15585/mmwr.mm6904a3] [Medline: 31999688]
2. Monto MA, McRee N, Deryck FS. Nonsuicidal self-injury among a representative sample of us adolescents, 2015. Am J Public Health 2018 Aug;108(8):1042-1048. [doi: 10.2105/AJPH.2018.304470] [Medline: 29927642]
3. Lo CB, Bridge JA, Shi J, Ludwig L, Stanley RM. Children's mental health emergency department visits: 2007-2016. Pediatrics 2020 Jun;145(6):e20191536 [FREE Full text] [doi: 10.1542/peds.2019-1536] [Medline: 32393605]

4. About Self-Injury. International Society for the Study of Self-Injury. 2020. URL: https://itriples.org/category/about-self-injury/ [accessed 2021-05-12]
5. Hawton K, Zahl D, Weatherall R. Suicide following deliberate self-harm: long-term follow-up of patients who presented to a general hospital. Br J Psychiatry 2003 Jun;182:537-542. [doi: 10.1192/bjp.182.6.537] [Medline: 12777346]
6. Whitlock J, Muehlenkamp J, Eckenrode J, Purington A, Baral Abrams G, Barreira P, et al. Nonsuicidal self-injury as a gateway to suicide in young adults. J Adolesc Health 2013 Apr;52(4):486-492. [doi: 10.1016/j.jadohealth.2012.09.010] [Medline: 23298982]
7. Moreno MA, Ton A, Selkie E, Evans Y. Secret society 123: understanding the language of self-harm on Instagram. J Adolesc Health 2016 Jan;58(1):78-84 [FREE Full text] [doi: 10.1016/j.jadohealth.2015.09.015] [Medline: 26707231]
8. Brown RC, Fischer T, Goldwich AD, Keller F, Young R, Plener PL. #cutting: non-suicidal self-injury (NSSI) on Instagram. Psychol Med 2018 Jan;48(2):337-346. [doi: 10.1017/S0033291717001751] [Medline: 28705261]
9. Brown RC, Bendig E, Fischer T, Goldwich AD, Baumeister H, Plener PL. Can acute suicidality be predicted by Instagram data? Results from qualitative and quantitative language analyses. PLoS One 2019;14(9):e0220623 [FREE Full text] [doi: 10.1371/journal.pone.0220623] [Medline: 31504042]
10. Lopez-Castroman J, Moulahi B, Azé J, Bringay S, Deninotti J, Guillaume S, et al. Mining social networks to improve suicide prevention: a scoping review. J Neurosci Res 2020 Apr 27;98(4):616-625. [doi: 10.1002/jnr.24404] [Medline: 30809836]
11. Wang Y, Tang J, Li J, Li B, Wan Y, Mellina C, et al. Understanding and Discovering Deliberate Self-harm Content in Social Media. In: Paper presented at the International World Wide Web Conference Committee. 2017 Apr 03 Presented at: IW3C2'17; June 2-5, 2017; Perth, Australia.
12. Coppersmith G, Leary R, Crutchley P, Fine A. Natural language processing of social media as screening for suicide risk. Biomed Inform Insights 2018 Aug 27;10:1178222618792860 [FREE Full text] [doi: 10.1177/1178222618792860] [Medline: 30158822]
13. Gomes de Andrade NN, Pawson D, Muriello D, Donahue L, Guadagno J. Ethics and artificial intelligence: suicide prevention on Facebook. Philos Technol 2018 Dec 26;31(4):669-684. [doi: 10.1007/s13347-018-0336-0]
14. Insel TR. Digital phenotyping: technology for a new science of behavior. J Am Med Assoc 2017 Oct 3;318(13):1215-1216. [doi: 10.1001/jama.2017.11295] [Medline: 28973224]
15. Jain SH, Powers BW, Hawkins JB, Brownstein JS. The digital phenotype. Nat Biotechnol 2015 May;33(5):462-463. [doi: 10.1038/nbt.3223] [Medline: 25965751]
16. Torous J, Kiang MV, Lorme J, Onnela J. New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research. JMIR Ment Health 2016;3(2):e16 [FREE Full text] [doi: 10.2196/mental.5165] [Medline: 27150677]
17. Guntuku SC, Yaden DB, Kern ML, Ungar LH, Eichstaedt JC. Detecting depression and mental illness on social media: an integrative review. Curr Opinion Behav Sci 2017 Dec;18:43-49. [doi: 10.1016/j.cobeha.2017.07.005]
18. Thorstad R, Wolff P. Predicting future mental illness from social media: a big-data approach. Behav Res Methods 2019 Aug;51(4):1586-1600 [FREE Full text] [doi: 10.3758/s13428-019-01235-z] [Medline: 31037606]
19. Kim J, Uddin ZA, Lee Y, Nasri F, Gill H, Subramanieapillai M, et al. A Systematic review of the validity of screening depression through Facebook, Twitter, Instagram, and Snapchat. J Affect Disord 2021 May 1;286:360-369. [doi: 10.1016/j.jad.2020.08.091] [Medline: 33691948]
20. Michelmore L, Hindley P. Help-seeking for suicidal thoughts and self-harm in young people: a systematic review. Suicide Life Threat Behav 2012 Oct;42(5):507-524. [doi: 10.1111/j.1943-278X.2012.00108.x] [Medline: 22889130]
21. Gerrard Y. Beyond the hashtag: circumventing content moderation on social media. New Media Soc 2018 May 28;20(12):4492-4511. [doi: 10.1177/1461444818776611]
22. Oldershaw A, Richards C, Simic M, Schmidt U. Parents' perspectives on adolescent self-harm: qualitative study. Br J Psychiatry 2008 Aug;193(2):140-144. [doi: 10.1192/bjp.bp.107.045930] [Medline: 18669999]
23. Murray CD, Warm A, Fox J. An internet survey of adolescent self-injurers. Aus J Adv Mental Health 2014 Dec 17;4(1):18-26. [doi: 10.5172/jamh.4.1.18]
24. Martinez-Martin N, Insel TR, Dagum P, Greely HT, Cho MK. Data mining for health: staking out the ethical territory of digital phenotyping. NPJ Digit Med 2018 Dec 19;1(1):68 [FREE Full text] [doi: 10.1038/s41746-018-0075-8] [Medline: 31211249]
25. Uhlhaas P, Torous J. Digital tools for youth mental health. NPJ Digit Med 2019 Oct 18;2(1):104-104 [FREE Full text] [doi: 10.1038/s41746-019-0181-2] [Medline: 31646184]
26. Radovic A, Badawy SM. Technology use for adolescent health and wellness. Pediatrics 2020 May;145(Suppl 2):S186-S194. [doi: 10.1542/peds.2019-2056G] [Medline: 32358210]
27. Auxier B, Rainie L, Anderson M, Perrin A, Kumar M, Turner E. Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal Information. Pew Research Center. 2019 Nov 15. URL: https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/ [accessed 2021-07-26]

28. Stoilova M, Nandagiri R, Livingstone S. Children’s understanding of personal data and privacy online – a systematic evidence mapping. Inform Commun Soc 2019 Sep 17;24(4):557-575. [doi: 10.1080/1369118x.2019.1657164]
29. Bietz MJ, Cheung C, Rubanovich CK, Schairer C, Bloss CS. Privacy perceptions and norms in youth and adults. Clin Pract Psychol 2019 Mar;7(1):93-103. [doi: 10.1037/cpp0000270]
30. Rooksby J, Morrison A, Murray-Rust D. Student Perspectives on Digital Phenotyping: The Acceptability of Using Smartphone Data to Assess Mental Health. In: CHI Conference on Human Factors in Computing Systems Proceedings. 2019 May 02 Presented at: CHI'19; January 23-26, 2019; Glasgow, Scotland UK.
31. Gulbas LE, Hausmann-Stabile C, De Luca SM, Tyler TR, Zayas LH. An exploratory study of nonsuicidal self-injury and suicidal behaviors in adolescent Latinas. Am J Orthopsychiatry 2015 Jul;85(4):302-314 [FREE Full text] [doi: 10.1037/ort0000073] [Medline: 26052816]
32. Claes L, Luyckx K, Bijttebier P. Non-suicidal self-injury in adolescents: Prevalence and associations with identity formation above and beyond depression. Personality and Individual Differences 2014 Apr;61-62:101-104. [doi: 10.1016/j.paid.2013.12.019]
33. Marshall SK, Tilton-Weaver LC, Stattin H. Non-suicidal self-injury and depressive symptoms during middle adolescence: a longitudinal analysis. J Youth Adolesc 2013 Aug;42(8):1234-1242. [doi: 10.1007/s10964-013-9919-3] [Medline: 23371004]
34. Kann L, McManus T, Harris WA, Shanklin SL, Flint KH, Queen B, et al. Youth Risk Behavior Surveillance - United States, 2017. MMWR Surveill Summ 2018 Dec 15;67(8):1-114 [FREE Full text] [doi: 10.15585/mmwr.ss6708a1] [Medline: 29902162]
35. Lewis-Fernández R, Das AK, Alfonso C, Weissman MM, Olfson M. Depression in US Hispanics: diagnostic and management considerations in family practice. J Am Board Fam Pract 2005;18(4):282-296 [FREE Full text] [doi: 10.3122/jabfm.18.4.282] [Medline: 15994474]
36. Barnett I, Torous J. Ethics, Transparency, and Public Health at the Intersection of Innovation and Facebook's Suicide Prevention Efforts. Ann Intern Med 2019 Apr 16;170(8):565-566. [doi: 10.7326/M19-0366] [Medline: 30743261]
37. Alegria M, Vallas M, Pumariega AJ. Racial and ethnic disparities in pediatric mental health. Child Adolesc Psychiatr Clin N Am 2010 Oct;19(4):759-774 [FREE Full text] [doi: 10.1016/j.chc.2010.07.001] [Medline: 21056345]
38. Birnholtz J, Burke M, Steele A. Untagging on social media: Who untags, what do they untag, and why? Computers in Human Behavior 2017 Apr;69:166-173. [doi: 10.1016/j.chb.2016.12.008]
39. Oeldorf-Hirsch A, Birnholtz J, Hancock JT. Your post is embarrassing me: Face threats, identity, and the audience on Facebook. Computers in Human Behavior 2017 Aug;73:92-99. [doi: 10.1016/j.chb.2017.03.030]
40. Marwick AE, boyd D. Networked privacy: How teenagers negotiate context in social media. New Media & Society 2014 Jul 21;16(7):1051-1067. [doi: 10.1177/1461444814543995]
41. America's Families and Living Arrangements. U.S. Census Bureau. 2016. URL: https://www.census.gov/data/tables/2016/demo/families/cps-2016.html [accessed 2021-05-21]
42. Vega WA, Rodriguez MA, Ang A. Addressing stigma of depression in Latino primary care patients. Gen Hosp Psychiatry 2010;32(2):182-191. [doi: 10.1016/j.genhosppsych.2009.10.008] [Medline: 20302993]
43. Latino Milwaukee: A Statistical Portrait. University of Wisconsin - Milwaukee, Center for Economic Development. 2016 Apr. URL: https://www.greatermilwaukeefoundation.org/files/7914/6215/2972/Latino_Milwaukee_Study_2016.pdf [accessed 2021-05-21]
44. Campos-Castillo C, Thomas B, Reyes F, Laestadius L. Seeking Help From Trusted Adults in Response to Peers’ Social Media Posts About Mental Health Struggles: Qualitative Interview Study Among Latinx Adolescents. JMIR Mental Health 2021;9(6) (forthcoming).
45. Whitehead R, Pringle J, Scott E, Milne D, McAteer J. The relationship between a trusted adult and adolescent health and education outcomes. Edinburg: NHS Health Scotland; 2019. URL: https://dera.ioe.ac.uk/34105/1/the-relationship-between-a-trusted-adult-and-adolescent-health-outcomes_6588.pdf [accessed 2021-05-21]
46. 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]
47. Barbour RS. Checklists for improving rigour in qualitative research: a case of the tail wagging the dog? BMJ 2001 May 5;322(7294):1115-1117 [FREE Full text] [Medline: 11337448]
48. Birks M, Chapman Y, Francis K. Memoing in qualitative research. Journal of Research in Nursing 2008 Jan;13(1):68-75. [doi: 10.1177/1744987107081254]
49. Berger E, Hasking P, Martin G. 'Listen to them': Adolescents' views on helping young people who self-injure. J Adolesc 2013 Oct;36(5):935-945. [doi: 10.1016/j.adolescence.2013.07.011] [Medline: 24011109]
50. Nissenbaum H. Privacy in Context. Stanford, CA: Stanford University Press; 2010:752367.
51. Bevan Jones R, Stallard P, Agha SS, Rice S, Werner-Seidler A, Stasiak K, et al. Practitioner review: Co-design of digital mental health technologies with children and young people. J Child Psychol Psychiatry 2020 Aug;61(8):928-940. [doi: 10.1111/jcpp.13258] [Medline: 32572961]
52. Facebook. Suicide Prevention. 2021. URL: https://www.facebook.com/safety/wellbeing/suicideprevention [accessed 2021-05-21]

53. TikTok. Suicide Prevention Resources. 2021. URL: https://www.tiktok.com/safety/resources/suicide-hotline-resources?lang=en&appLaunch= [accessed 2021-05-21]
54. Abelson R, Goodnough A. If the Supreme Court Ends Obamacare, Here's What It Would Mean. The New York Times. 2021 Mar 22. URL: https://www.nytimes.com/article/supreme-court-obamacare-case.html [accessed 2021-05-21]
55. Schneble CO, Elger BS, Shaw DM. All Our Data Will Be Health Data One Day: The Need for Universal Data Protection and Comprehensive Consent. J Med Internet Res 2020 May 28;22(5):e16879. [doi: 10.2196/16879]
56. Cummings JR, Druss BG. Racial/ethnic differences in mental health service use among adolescents with major depression. J Am Acad Child Adolesc Psychiatry 2011 Feb;50(2):160-170 [FREE Full text] [doi: 10.1016/j.jaac.2010.11.004] [Medline: 21241953]
57. Anderson M, Jiang J. Teens, Social Media, and Technology 2018. Pew Research Center. 2018 May 31. URL: http://www.pewinternet.org/2018/05/31/teens-social-media-technology-2018/ [accessed 2021-05-21]
58. Ault-Brutus A, Alegria M. Racial/ethnic differences in perceived need for mental health care and disparities in use of care among those with perceived need in 1990-1992 and 2001-2003. Ethn Health 2018 Feb;23(2):142-157. [doi: 10.1080/13557858.2016.1252834] [Medline: 27809570]
59. Mechanic D. Disadvantage, inequality, and social policy. Health Aff (Millwood) 2002;21(2):48-59. [doi: 10.1377/hlthaff.21.2.48] [Medline: 11900186]
60. Phelan JC, Link BG, Tehranifar P. Social conditions as fundamental causes of health inequalities: theory, evidence, and policy implications. J Health Soc Behav 2010;51 Suppl:S28-S40. [doi: 10.1177/0022146510383498] [Medline: 20943581]
61. Pretorius C, Chambers D, Coyle D. Young People's Online Help-Seeking and Mental Health Difficulties: Systematic Narrative Review. J Med Internet Res 2019 Nov 19;21(11):e13873 [FREE Full text] [doi: 10.2196/13873] [Medline: 31742562]
62. Frost M, Casey L. Who Seeks Help Online for Self-Injury? Arch Suicide Res 2016;20(1):69-79. [doi: 10.1080/13811118.2015.1004470] [Medline: 25706352]

Abbreviations

NSSI: nonsuicidal self-injury

Edited by R Kukafka; submitted 18.03.21; peer-reviewed by Y Evans, C Biernesser, K Braune; comments to author 16.04.21; revised version received 22.04.21; accepted 05.07.21; published 10.08.21.

Please cite as:
Laestadius LI, Craig KA, Campos-Castillo C
Perceptions of Alerts Issued by Social Media Platforms in Response to Self-injury Posts Among Latinx Adolescents: Qualitative Analysis
J Med Internet Res 2021;23(8):e28931
URL: https://www.jmir.org/2021/8/e28931
doi:10.2196/28931
PMID:34383683

©Linnea I Laestadius, Katherine A Craig, Celeste Campos-Castillo. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Psychological Violence Against Arab Women in the Context of Social Media: Web-Based Questionnaire Study

Ahmed Omar Bali[1*], PhD; Emad Omer[2*], PhD; Kawa Abdulridha[3*], PhD; Araz Ramazan Ahmad[4,5*], PhD

[1] Diplomacy and Public Relations Department, University of Human Development, Sulaimani, Iraq
[2] Mass Communication Department, Ajman University, Ajman, United Arab Emirates
[3] Sulaimani Polytechnic University, Sulaymaniyah, Iraq
[4] Department of Administration, College of Humanities, University of Raparin, Ranya, Iraq
[5] Department of International Relations & Diplomacy, Faculty of Administrative Sciences and Economics, Tishk International University, Erbil, Iraq
[*] all authors contributed equally

Corresponding Author:
Ahmed Omar Bali, PhD
Diplomacy and Public Relations Department
University of Human Development
15 Dania
Sulaimani, SD10A
Iraq
Phone: 964 7736990915
Email: ahmed.bali@uhd.edu.iq

Abstract

Background: Social media provides women with varying platforms to express themselves, show their talents, communicate and expand their social relationships, and break the shackles imposed by their societies. Theoretically, social media can play a significant role in developing women’s freedom and decreasing social pressures; nonetheless, women continue to face violence during the social media era mainly in the form of psychological violence.

Objective: This study aims to conduct an empirical in-depth analysis of how the digital space, particularly social media, provides men with new opportunities to surveil, restrict, harass, and intimidate feminists in Arab countries.

Methods: This study includes an empirical survey to investigate what Arab women think are the causes and types of violence wielded against them and their perspectives on the impact of that violence. This study used a web-based questionnaire administered through Google Forms (n=1312) with responses from Arab women aged 15 years and above from all Arab countries.

Results: We found that most Arab women feared posting an actual photograph of themselves on their social media accounts and only approximately one-third (490/1312, 37.3%) did so. Most women indicated that they encountered sexual harassment regardless of their age. Furthermore, most women were not aware of the legal aspects of this crime and even those who were aware indicated that they would not press charges for several reasons, including bringing dishonor upon their families, the time-consuming nature of litigation, and fear of revenge.

Conclusions: This study shows that young and less educated women are more vulnerable to abuse from either social media users or being condemned by their families. This has several effects, including lower self-esteem and hesitancy in seeking a job, feelings of mistrust and fear, cynicism, anxiety, depression, and sleep disorders. These issues hold women back from using social media in positive ways and some consider leaving social media.

(J Med Internet Res 2021;23(8):e27944) doi:10.2196/27944

KEYWORDS

psychological; violence; Arab women; social media; feminism; sociology; abuse; oppression; self-esteem

Introduction

Background

Social media provides women with a variety of platforms to express themselves [1-3] and show their talents, communicate, expand their social relationships, and break the shackles imposed by their societies. Theoretically, social media can play a significant role in developing women's freedom and decreasing social pressures on them [4]; nonetheless, women face high levels of violence on social media, particularly sexual harassment [5-9]. According to an Amnesty International report in 2018 [10], women are subjected to scorn, disrespect, and abuse on the internet, and they encounter a staggering amount of violent or aggressive behavior from their relatives and other individuals. Women in Arab countries use social media with fear and extreme caution. The behavior on the part of others that elicits this fear can be considered violence against women. Moreover, it can be seen as restricting freedom of speech and freedom of thought because of social media's encompassing role that extends beyond entertainment and leisure and into work and political life as well. Therefore, any kind of violence on social media directed against women negatively affects the physical, mental, financial, and social aspects of their lives.

As a phenomenon, violence against women on social media has not been studied sufficiently to understand this kind of violence better, the reasons for it, and the outcomes of that violence across the Arab region, with the majority of studies focusing on individual countries. This study aims to conduct an empirical in-depth analysis to investigate how the digital space negatively impacts women in Arab countries, using questionnaire surveys. We speculate that social media provides men with new opportunities to surveil, restrict, harass, and intimidate women in Arab countries.

Conceptual Framework

The intersection of women, feminism, and technology is addressed in many social science disciplines, particularly media and communications political science, sociology, and philosophy [9,11,12]. In order to examine Arab women's relationship with technology, particularly social media, which theoretically deliberates and empowers women, the dominant feminist approaches must be outlined. However, in this study, the liberal feminist approach will be highlighted to focus on the aspect of women's self-expression. The conceptual framework considers violence against women across social media, a phenomenon that is more emphasized by radical feminists. Besides the two approaches, *liberal feminist* and *radical feminist*, related studies in Arab countries will be considered.

Liberal Feminism and Social Media

The internet, particularly social media, is highly connected to the liberal feminist perspective and ambition. In this context, Megarry argues that "liberal feminism attempts to harness the liberal values of justice, equality, and fairness to fight for women's rights within the patriarchal state system" [9]. The liberal approach looks positively on the role played by technology, including social media, as a way to enhance women's lives. Although women face violence across digital technologies, they argue that social media can still bolster women's interests and rights [13]. The connection between social media and feminism has been examined in several interdisciplinary studies [14-19], in which researchers highlighted the positive aspects of social media that allow women to engage globally [16], contribute to social movements that reflect feminist goals [20], and expand and raise awareness about issues related to sexism, misogyny [21], inequality, and gendered violence [17]. In a similar context, Crossley argues that social media facilitates understanding of women's issues and enhances feminist ideas across societies [18]. Facebook, Twitter, Snapchat, and Instagram are described as interactive and participatory platforms, which allow for the dissemination of information through the creation, sharing, and consumption of text, audio, and video content [22-24]. From this standpoint, social media enables feminist activists and groups to build and create interactive networks to achieve their goals and shape positive ideas about women and their interests in the face of beliefs and ideas that oppose equality for women.

Radical Feminism and Social Media

The radical feminist perspective tends to view the digital space more negatively, seeing women as exposed to abuse and domination of the male perspective [25-27]. Social media enhances radical feminism, which ideologically becomes fragmented and less geographically bound [28]. In other words, radical feminism as a universal approach has been negatively affected because of social media. Even before the emergence of social media, the radical feminist approach rejected the argument that the developments in technology resulted in positive outcomes for women [29-32] because they believe that social institutions are dominated by men [33] and new technology, including social media, is developed in the interest of men [9]. Many studies emphasize this dark side of social media with relation to women. In this regard, Stubbs-Richardson et al [14] show that women are often pathologized or trolled either on the internet or offline and cyberspace is a sexist and gendered space [14] controlled by the logic of power, ideology, and a market-oriented approach. Other studies argue that social media as a commercial platform aims to generate revenue [34]. Other studies, focused on Tinder and dating apps, show that gendered discourses have implications on hidden harassment and abuse of women [35]. From the radical feminist perspective, women face violence and are pathologized or trolled and exploited by men.

Status of Arab Women

Like all women, Arab women are affected by cultural factors such as religious beliefs, social systems, education, and the media [36]. Religious beliefs play a remarkable role in social and cultural socialization, working to determine women's relationship with others, even within the realm of social media [37]. The education system has also portrayed Arab women negatively, and they are directed to deal with men sensitively and with fear [38]. In turn, this results in a tendency for them to adopt conservative values because they still follow the traditional Arab culture where their relationships are limited. While liberal feminists might argue that, in the age of satellite television channels there is opportunity for openness, Arab media tends to focus on women's bodies as a commodity and promotes women as key to the labor market and in constructing the family and society. Other studies have found that social media has worked to expand the social circles of Arab women [39]. However, they tend to face high levels of violence, with the Arab League releasing a report that Arab women tend to be faced with a lack of real political participation, education, job opportunities, and heath care and are vulnerable to violence as a consequence of regional war and conflict [40]. Similarly, the Economic and Social Commission for Western Asia has released reports that corroborate the Arab League's findings [40]. Overall, there is a lack of studies on violence on social media against Arab women, with these 2 reports failing to address the issue. Hence, this study seeks to examine this issue to understand the present dynamics and identify possible solutions in a better manner.

Research Questions and Hypotheses

On the basis of this literature review and our study objectives, 4 questions and 8 hypotheses were developed.

Research Questions

1. How do Arab women use social media?
2. Do Arab women use social media freely?
3. How do Arab women face sexual harassment via social media?
4. Do Arab women have legal awareness about sexual harassment on social media?

Hypotheses

Along with the research questions shown above, this study tests 8 hypotheses to shed light on the study's underlying arguments.

1. Level of education is positively associated with posting personal photographs on social media accounts.
2. Age is negatively associated with posting personal photographs on social media accounts.
3. Level of education is negatively associated with being condemned for posting a photograph or about a topic on social media.
4. Age is negatively associated with being condemned for posting a photograph or about a topic on social media.
5. Level of education is associated with being sexually harassed via social media.
6. Age is not associated with being sexually harassed via social media.
7. Level of education is associated with legal awareness about sexual harassment on social media.
8. Level of education is associated with willingness to press charges against sexual harassers.

Methods

Methods Overview

This study includes an empirical survey to investigate what Arab women think of the types and causes of violence against them and their perspectives about the impact of that violence. This study used a web-based questionnaire administered through Google Forms (n=1312) with responses from Arab women aged 15 years and older from all Arab countries. The random sampling method is considered one of the best methods since everyone in the research context had an opportunity to participate [41,42]. Consequently, the results represented the population as a whole [43]. Data collection was carried out starting during the last week of October through mid-December 2020. The participants were categorized into five age groups: 15-18 years (134/1312, 10.2%); 19-25 years (608/1312, 46.3%), which was the largest group; 26-35 years (250/1312, 19.1%); 36-50 years (264/1312, 20.1%); and ≥51 years (56/1312, 4.3%). The 19-25–year-old participants comprised the largest group because they are mainly college students, a cohort that actively uses social media. The survey also considered the marital status of the participants. Single women comprised the largest number with 858 (of 1312, 65.4%) participants, married women comprised the second-largest group with 414 (of 1312, 31.6%) participants, and widowed/divorced women comprised the smallest group with 40 (of 1312, 3.0%) participants. The participants were from different educational levels: those who only have reading and writing skills (8/1312, 0.6%), high school students (138/1312, 10.5%), university students (454/1312, 34.6%), those with a bachelor's degree (446/1312, 34.0%), those with a master's or PhD degree (244/1312, 18.6%), and activists (22/1312, 1.7%)—referring to those working in political parties and organizations. The demographic characteristics of the study participants are summarized in Table 1. A Spearman correlation model was used to test the aforementioned 8 hypotheses.

Table 1. Demographic backgrounds of women (N=1312).

Characteristics	Women, n (%)
Age (years)	
15-18	134 (10.2)
19-25	608 (46.3)
26-35	250 (19.1)
36-50	264 (20.1)
≥51	56 (4.3)
Education levels	
Reading and writing skills	8 (0.6)
High school	138 (10.5)
University students	454 (34.6)
Bachelor's degree	446 (34.0)
Master's or PhD degree	244 (18.6)
Activists	22 (1.7)
Marital status	
Single	858 (65.4)
Married	414 (31.6)
Widowed/divorced	40 (3.0)

Procedure for Measuring Hypotheses

The data were analyzed by using SPSS (version 22, IBM Corp). Considering the nature and aim of this study, several questions were presented in the questionnaire to measure psychological violence against Arab women through binary scale ("yes/no") and triple scale ("yes/no/do not know") modes. "Do not know" was used for some questions, which we predicted that some of the participants would answer with "do not know" to express their views. Furthermore, a triple scale mode "no/somewhat/yes" was used. The "somewhat" response was included to know how many women face sexual harassment or whether they face it rarely. The second model was a Spearman correlation model with Sig. (2-tailed) *P* values, which was processed in SPSS to test the 8 hypotheses through 2 levels of public engagement.

Results

Social Media Use by Arab Women

This section examines the first research question, which asked the women who participated to identify their preferred social media platform. Table 2 shows that the majority (770/1312, 58.7%) prefer Facebook and one-fifth (272/1312, 20.7%) prefer Instagram. The third-most popular platform among Arab women is Snapchat, which was used by 128 (9.8%) participants. As shown in Table 3, the user-friendliness of a social media platform was most important to them, which 808 (61.6%) participants selected as the most important reason they chose their preferred platform. The second-most cited reason was cybersecurity, chosen by 138 (10.5%) participants. This reason is important for Arab women, but the majority of them did not highlight this issue because hacking a social media account is not easy and women tend to use social media cautiously (Table 3). In total, 122 (9.3%) Arab women use social media for promotions and personal branding.

Table 2. Popular social media platform among Arab women (N=1312).

Social media platforms	Women, n (%)
Facebook	770 (58.7)
Instagram	272 (20.7)
Snapchat	128 (9.8)
TikTok	94 (7.2)
WhatsApp	48 (3.7)

Table 3. Reasons for using such social media platforms among Arab women (N=1312).

Reasons	Women, n (%)
Cybersecurity	138 (10.5)
User-friendly	808 (61.6)
Most popular	124 (9.5)
Entertainment	120 (9.1)
Useful for promotion and personal branding	122 (9.3)

Freedom of Using Social Media by Arab Women

This section examines the second research question, which looks at the freedom with which Arab women use social media. Table 4 shows that most women are cautious about posting photographs of themselves on their social media accounts and only approximately one-third (490/1312, 37.3%) do so. Approximately half (628/1312, 47.9%) of respondents post general photographs, such as those of a public figure, an image of nature, a city, or text with an image instead. Thus, they feel that they cannot post a photograph of themselves, but still seek to represent their personal identity. Table 4 shows that 180 (13.7%) participants post group photographs of family members that include themselves, which indicates that they would like to post personal photographs but are under pressure not to do so alone. As shown in Table 5, the main reason expressed by the 726 women who did not post their personal photograph on their social media accounts was fear that it would be misused (488/1312, 67.2%). This represents a major issue for women because they feel that it can affect their reputation. Table 5 shows that 112 (15.4%) respondents claim that they feel ashamed to show their photograph because Arab society is generally conservative, and women cannot express themselves even on social media. Seventy-four (10.1%) women revealed that their families do not allow them to post their personal photographs on social media. Their families fear the misuse of the photograph, which can affect reputation and honor.

Table 4. Types of images on social media accounts of Arab women (N=1312).

Type of image	Women, n (%)
Personal photograph	490 (37.3)
Photograph that contains me and my family	180 (13.7)
Photograph that contains me and my friends	14 (1.1)
General images such as those of a public figure, nature, a city, or a text with an image	628 (47.9)

Table 5. Reasons for not posting personal photographs on social media accounts.

Reasons	Women, n (%)
Fear of misusing my photograph by other people	488 (67.2)
My family does not allow me to post my photograph on my social media account	74 (10.1)
Society and cultural pressure	52 (7.1)
Feeling shameful	112 (15.4)
Total	726 (100)

Sexual Harassment via Social Media

This section examines the third research question, which addresses how Arab women face sexual harassment via social media. As shown in Table 6, at total of 494 (37.7%) participants stated that they encountered sexual harassment, and 508 (38.7%) stated that they somewhat encountered sexual harassment. Less than one-fourth (310/1312, 23.6%) of respondents claimed that they have not been sexually harassed via social media (Table 6). Thus, men are given yet another opportunity to engage in violence against women, beyond the public setting, workplace, and at home. Moreover, sexual harassment is often a disturbing experience. Survivors can be subjected to significant harm that not only inflicts emotional distress but also raises substantial problems in their family situations.

Table 6. Sexual harassment of Arab women via social media (N=1312).

Sexually harassed via social media	Women, n (%)
Yes	494 (37.7)
Somewhat	508 (38.7)
No	310 (23.6)

Women's Legal Awareness of Sexual Harassment on Social Media

This section examines the fourth research question, which addresses legal awareness about sexual harassment on social media. Table 7 indicates that only half of the participants (646/1298, 49.8%) were aware that charges could be pressed against sexual harassment on the internet, and 222 (17.1%) were unaware. Approximately all 1264 (98.3%) participants would not take legal action against sexual harassers. As shown in Table 8, half (642/1298, 49.9%) of the participants did not want to take legal action against sexual harassers because they believe that it could affect their reputation. Another reason stated by many women (373/1298, 29%) was that they thought that the process of a sexual harassment lawsuit would take too long. Another 271 (21.1%) participants feared acts of revenge by sexual harassers. These findings indicate that women do not have complete confidence in judicial power and the police to protect them.

Table 7. Perceptions of sexual harassment as a punishable offense and willingness to press charges against harassers (N=1298).

Perception of sexual harassment being a punishable offence	Women, n (%)
Do you think sexual harassers can be punished?	
Yes	646 (49.8)
No	222 (17.1)
Do not know	646 (49.8)
Would you take sexual harassers to court ?	
Yes	22 (1.7)
No	1264 (98.3)

Table 8. Reasons of not pressing charges against sexual harassers among Arab women (N=1286).

Reasons	Women, n (%)
It takes my time in the court	373 (29.0)
It affects my reputation	642 (49.9)
I fear that the accused will take revenge	271 (21.1)

Results of Our Study Hypotheses

This section is focused on testing and interpreting the 8 study hypotheses. The first hypothesis predicted a positive correlation between the level of education of Arab women and posting their personal photograph on social media. Table 9 shows that this hypothesis (R_s=–0.085; P<.001) is negatively correlated, which implies that women with lower education levels were more likely than those with higher education levels to post their personal photograph on social media.

The second hypothesis predicted that older Arab women were more likely to post the personal photograph on their social media accounts than younger women. This hypothesis (R_s=–0.134; P<.001) was negatively correlated, similar to the first hypothesis. We assumed that educated and older women would understand and recognize problems such as sexual harassment better than relatively lesser educated and younger women and would therefore be more reluctant to post their photograph on social media.

The third hypothesis predicted that Arab women with higher levels of education would be less likely to face pressure for posting a photograph or a topic on social media than those who are relatively less educated because it is a common prediction that education empowers women to become more independent and respected among families, which this present study confirmed (R_s=–0.076; P=.02).

The fourth hypothesis predicted that age is negatively associated with being condemned for posting a photograph or topic on social media. This hypothesis (R_s=–0.053; P=.06) is not confirmed, which implies that Arab women are more likely being condemned for posting a photograph or topic on social media regardless of age.

The fifth hypothesis predicted a positive correlation between the education level of women and being sexually harassed via social media. This hypothesis (R_s=0.083; P=.04) was confirmed, which implies that educated women face more sexual harassment via social media. It is assumed that they are more active in the public sphere, have wider relationships, and are more popular than lesser educated women, the latter probably spending more time within their families.

The sixth hypothesis predicted that age is negatively associated with being sexually harassed via social media, which implies that younger women more face sexual harassment. However, Table 9 shows that this hypothesis (R_s=0.036; P=.003) was not confirmed, which implies that Arab women are likely to face sexual harassment on social media irrespective of their age. Table 9 shows that 494 (37.7%) participants stated that they encountered sexual harassment, and 508 (38.7 %) stated that they somewhat encountered it.

The seventh hypothesis predicted that the education level of women is associated with legal awareness about sexual harassment on social media, and this hypothesis (R_s=0-125; *P*<.001) was confirmed.

The eighth hypothesis predicted that the education level of women is associated with a willingness to press charges against sexual harassers. This hypothesis predicted that educated women have more power and realize that the sexual harassment can be legally prosecuted and will be more likely to confront their harassers in court. However, this hypothesis (R_s=0.014; *P*=.62) was not confirmed because the vast majority of participants (1264/1286, 98.3%) were not willing to press charges against their sexual harassers irrespective of their educational background (Table 7).

Table 9. The Spearman correlation model reports the associations between age and education levels of Arab women with posting their personal photographs on social media, being condemned for using social media, experiencing sexual harassment, and having legal awareness of sexual harassment on social media (N=1312).

Hypothesis	Variables	R_s value	Sig. (2-tailed) *P* value
1	Level of education is positively associated with posting a personal photograph on social media.	–0.085[a]	.002
2	Age is associated negatively with posting a personal photograph on social media.	–0.134[a]	<.001
3	Level of education is negatively associated with being condemned for posting a photograph or topic on social media.	–0.076[a]	.02
4	Age is negatively associated negatively with being condemned for posting a photograph or topic on social media.	–0.053	.06
5	Education level of women is positively associated with being sexually harassed via social media.	0.083[a]	.06
6	Age of women is not associated with being sexually harassed via social media.	0.036	.19
7	Education level of women is associated with legal awareness of sexual harassment on social media.	0.125[a]	<.001
8	Education level of women is associated with willingness to pressing charges against sexual harassers.	0.014	.62

[a] The correlation is significant when R_s values range from –1 to +1.

Discussion

Principal Findings

There is a lack of studies examining the positive and negative aspects of social media in the lives of Arab women [39]. Previous studies at the global level have referred to a number of positive aspects of social media platforms, such as their role in facilitating interaction and participation and disseminating information [22-24]. Other studies have highlighted the positive aspects of social media in boosting the feminist movement [20], understanding women's issues [18], and raising awareness about sexism and misogyny [21]. In this study, only 122 (9.3%) participants indicated that they used social media for promotions and personal branding (Table 3) [44,45]. These positive aspects of social media help women to become entrepreneurs and launch web-based businesses and other positive ventures. Besides this, 120 (9.1%) participants indicated that social media is an entertainment platform for them and social media lets them access global culture, even if they live in a relatively conservative society (Table 3). Fear of image-based abuse, blackmailing, and interference in their private lives were reasons that motivated half of the participants to refrain from posting personal photographs on their social media profiles. For instance, Table 4 indicates that most Arab women were reluctant to post their personal photographs on their social media accounts and only approximately one-third (490/1312, 37.3%) did so. Alternatively, they posted images of local or international celebrities, nature, or family group photographs, which are more difficult to alter or use maliciously. This situation counters the liberal notion that women can express themselves on social media and use social media as a platform to play influential roles in society [1-3,16]. Even though laws in many Arab countries clearly define harassment on social media as a crime, many women encounter some level of sexual harassment (Table 6) irrespective of their age. Although some women knew that this is a crime (Table 7), they indicated that they would not come forward and press charges against their harassers for several reasons, including bringing dishonor upon their families, the time-consuming nature of litigation, and the fear of revenge.

Limitations

The majority of previous studies have focused on individual Arab countries rather than adopting a regional approach, and have used descriptive rather than data-based approaches. This does not allow for extensive comparison with this study's outcomes. Regarding data collection, doing so in-person across all Arab countries representationally is difficult and costly. This was also prevented by the COVID-19 pandemic in 2020. Therefore, this study adopted a survey method using a web-based questionnaire administered through Google Forms (N=1312) with responses from Arab women aged ≥15 years from all Arab countries. More extensive data collection would further improve our results.

Conclusions

This study predicts that women's endurance of the vast majority of sexual harassment without taking legal action against sexual harassers could motivate harassers to continue engaging in abusive behavior. This harms women and creates a climate of intimidation and repression in social and economic life, particularly among young women, as this study shows that age is negatively associated with being condemned for posting a photograph or topic on social media (Table 9). Conversely, the education level of women plays a positive role in independence within the family. Our third hypothesis predicts that Arab women with lower levels of education might face more pressure for posting a photograph or topic on social media than those with higher education levels. This study found that young women and the relatively lesser educated are more likely to face abuse. This leads to several issues including lower self-esteem and hesitancy in looking for a job, feelings mistrust and fear, cynicism, anxiety, depression, and sleep disorders. These issues constrain women from using social media and some of them think about disengaging from social media. This issue needs governments' consideration to develop laws to reduce violence and empower women to take legal action. Hence, future studies should address the violence against women with regard to legal and psychological perspectives.

Acknowledgments

This study was funded by Ajman University, United Arab Emirates.

Conflicts of Interest

None declared.

References

1. Alsudairy HT. The Role of the Social Media in Empowering Saudi Women's Expression. Newcastle upon Tyne: Cambridge Scholars Publishing; 2020.
2. Andreassen CS, Pallesen S, Griffiths MD. The relationship between addictive use of social media, narcissism, and self-esteem: Findings from a large national survey. Addict Behav 2017 Jan;64:287-293. [doi: 10.1016/j.addbeh.2016.03.006] [Medline: 27072491]
3. Howley K. Understanding Community Media. Thousand Oaks, CA: Sage Publications; 2009.
4. Hagerty T, Hill T. Finding Peace in Iraq Joint Field Research on New Approaches to Peacebuilding in the Kurdistan Region. Center for Global Affairs. 2013. URL: https://www.sps.nyu.edu/content/dam/sps/academics/departments/center-for-global-affairs/initiatives/pdf-files/CGA-IPE-Finding-Peace-in-Iraq-Vol-1.pdf [accessed 2021-07-29]
5. Peuchaud S. Social media activism and Egyptians' use of social media to combat sexual violence: an HiAP case study. Health Promot Int 2014 Jun;29 Suppl 1:i113-i120. [doi: 10.1093/heapro/dau046] [Medline: 25217347]
6. Won D, Steinert-Threlkeld ZC, Joo J. Protest Activity Detection and Perceived Violence Estimation from Social Media Images. In: Proceedings of the 25th ACM international conference on Multimedia. 2017 Presented at: 25th ACM international conference on Multimedia; October 23-27, 2017; Mountain View, CA p. 786-794. [doi: 10.1145/3123266.3123282]
7. Chang E. Brotopia: Breaking Up the Boys' Club of Silicon Valley. New York, NY: Penguin; 2018.
8. Wiener A. Uncanny Valley: A Memoir. New York, NY: MCD Books; 2020.
9. Megarry J. Unravelling the Web of Equals. In: The Limitations of Social Media Feminism. Cham: Palgrave Macmillan; 2020.
10. Amnesty International Report 2017/18. The State of the World's Human Rights. Amnesty International. 2018. URL: https://www.amnesty.org/download/Documents/POL1067002018ENGLISH.PDF [accessed 2021-08-04]
11. Wajcman J. Feminist theories of technology. Camb J Econ 2009 Jan 08;34(1):143-152. [doi: 10.1093/cje/ben057]
12. Wyatt S. Feminism, technology and the information society: Learning from the past, imagining the future. Inf Commun Soc 2008 Feb;11(1):111-130. [doi: 10.1080/13691180701859065]
13. Halbert D. Shulamith Firestone. Inf Commun Soc 2004 Jan;7(1):115-135. [doi: 10.1080/1369118042000208933]
14. Stubbs-Richardson M, Rader NE, Cosby AG. Tweeting rape culture: Examining portrayals of victim blaming in discussions of sexual assault cases on Twitter. Fem Psychol 2018 Feb 08;28(1):90-108. [doi: 10.1177/0959353517715874]
15. Locke A, Lawthom R, Lyons A. Social media platforms as complex and contradictory spaces for feminisms: Visibility, opportunity, power, resistance and activism. Fem Psychol 2018 Feb 08;28(1):3-10. [doi: 10.1177/0959353517753973]
16. Guillard J. Is feminism trending? Pedagogical approaches to countering (Sl)activism. Gender and Education 2016 Jan 22;28(5):609-626. [doi: 10.1080/09540253.2015.1123227]
17. Baer H. Redoing feminism: digital activism, body politics, and neoliberalism. Fem 2015 Oct 16;16(1):17-34. [doi: 10.1080/14680777.2015.1093070]
18. Crossley A. Facebook Feminism: Social Media, Blogs, and New Technologies of Contemporary U.S. Feminism. Mobilization 2015;20(2):253. [doi: 10.17813/1086-671X-20-2-253]
19. Manago AM. Negotiating a sexy masculinity on social networking sites. Fem Psychol 2013 May 24;23(4):478-497. [doi: 10.1177/0959353513487549]
20. Earl J, Kimport K. Digitally Enabled Social Change: Activism in the Internet Age. Cambridge, MA: MIT Press; 2011.
21. Turley E, Fisher J. Tweeting back while shouting back: Social media and feminist activism. Fem Psychol 2018 Feb 08;28(1):128-132. [doi: 10.1177/0959353517715875]

22. Lyons A, McCreanor T, Goodwin I, Barnes HM. Youth Drinking Cultures in a Digital World: Alcohol, social media and cultures of intoxication. London: Routledge; 2017.
23. Carah N. Alcohol corporations and marketing in social media. In: Youth Drinking Cultures in a Digital World. London: Routledge; 2017:115-131.
24. Locke A, Lawthom R, Lyons A. Social media platforms as complex and contradictory spaces for feminisms: Visibility, opportunity, power, resistance and activism. Fem Psychol 2018 Feb 08;28(1):3-10. [doi: 10.1177/0959353517753973]
25. Citron DK. Hate Crimes in Cyberspace. Cambridge, MA: Harvard University Press; 2014.
26. Jane EA. Misogyny Online: A Short (and Brutish) History. Thousand Oaks, CA: Sage Publications; 2016.
27. Megarry J. Online incivility or sexual harassment? Conceptualising women's experiences in the digital age. Womens Stud Int Forum 2014 Nov;47:46-55. [doi: 10.1016/j.wsif.2014.07.012]
28. Maddison S, Frances S. Feminist Perspectives on Social Movement Research. In: Handbook of Feminist Research: Theory and Praxis. Thousand Oaks, CA: Sage Publications; 2012:413-433.
29. Arditti R, Klein R, Minden S. Test-tube Women: What Future for Motherhood?. Kitchener, ON: Pandora Press; 1984.
30. Corea G. The Mother Machine: Reproductive technologies from Artificial Insemination to Artificial Wombs. Am J Matern Child Nurs 1986;11(5):357-363. [doi: 10.1097/00005721-198609000-00014]
31. Fox MF, Johnson DG, Rosser SV. Genetic Technology and Women. In: Women, Gender, and Technology. Champaign, IL: University of Illinois Press; 2006:111-121.
32. Burfoot A. Women as wombs: Reproductive technologies and the battle over women's freedom. Womens Stud Int Forum 1994 Sep;17(5):552-553. [doi: 10.1016/0277-5395(94)90057-4]
33. Bell D, Klein R. Radically Speaking: Feminism Reclaimed. North Geelong: Spinifex Press; 1996.
34. Uhl CA, Rhyner KJ, Terrance CA, Lugo NR. An examination of nonconsensual pornography websites. Fem Psychol 2018 Feb 08;28(1):50-68. [doi: 10.1177/0959353517720225]
35. Thompson L. “I can be your Tinder nightmare”: Harassment and misogyny in the online sexual marketplace. Fem Psychol 2018 Feb 08;28(1):69-89. [doi: 10.1177/0959353517720226]
36. Aloud N, Rathur A. Factors Affecting Attitudes Toward Seeking and Using Formal Mental Health and Psychological Services Among Arab Muslim Populations. J Muslim Mental Health 2009 Oct 30;4(2):79-103. [doi: 10.1080/15564900802487675]
37. Farahat AM. Challenges and opportunities granted to Arab women by social media during the political upheavals between 2011 and 2018 (Master's Thesis)). Kadir Has University. 2020. URL: http://academicrepository.khas.edu.tr/bitstream/handle/20.500.12469/3936/ASMAA%20MEDHAT%20FARAHAT.pdf?sequence=1&isAllowed=y [accessed 2021-08-05]
38. Ty R, Alkarzon M, Alkarzon A. Arab Women, Education, Policy, and Empowerment: The Role of Gender, Culture, Religion, Traditionalism, and Class. J Modern Educ Rev 2015 Jan 20;5(1):49-58 [FREE Full text] [doi: 10.15341/jmer(2155-7993)/01.05.2015/005]
39. Odine M. Role of Social Media in the Empowerment of Arab Women. Glob Media J 2013:1-30 [FREE Full text]
40. Violence against women online in 2018. Amnesty International. URL: https://www.amnesty.org/en/latest/research/2018/12/rights-today-2018-violence-against-women-online/ [accessed 2021-07-29]
41. Becker S, Bryman A, Ferguson H. Understanding research for social policy and social work (second edition): Themes, methods and approaches. Bristol: Bristol University Press; 2012.
42. Lampard R, Pole C. Practical Social Investigation: Qualitative and Quantitative Methods in Social Research. London: Routledge; 2015.
43. Graue C. Qualitative data analysis. Int J Retail Marketing 2015;4(9):5 [FREE Full text]
44. Bali AO, Jabar S, Jalal H, Sofi-Karim M. Iraqi media entrepreneurs across social media: Factors and challenges. Journal of Digital Media & Policy 2021. [doi: 10.1386/jdmp_00024_1]
45. Bali A, Zarea H. The Challenges of Firms in Iraqi Kurdistan Economy in the Light of Strategic Acquisition Theory. In: Competitiveness in Emerging Markets. Cham: Springer; 2018:245-262.

Edited by R Kukafka; submitted 14.02.21; peer-reviewed by MA Bahrami, A Naser, L Schwab-Reese; comments to author 17.04.21; revised version received 01.07.21; accepted 05.07.21; published 19.08.21.

Please cite as:
Omar Bali A, Omer E, Abdulridha K, Ahmad AR
Psychological Violence Against Arab Women in the Context of Social Media: Web-Based Questionnaire Study
J Med Internet Res 2021;23(8):e27944
URL: https://www.jmir.org/2021/8/e27944
doi:10.2196/27944
PMID:34420919

©Ahmed Omar Bali, Emad Omer, Kawa Abdulridha, Araz Ramazan Ahmad. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Association Between Social Media Use and Cancer Screening Awareness and Behavior for People Without a Cancer Diagnosis: Matched Cohort Study

Lei Qin[1*], PhD; Xiaomei Zhang[1], MS; Anlin Wu[2], MS; James S Miser[3,4,5], MD; Yen-Lin Liu[4,6,7], MD, PhD; Jason C Hsu[8], PhD; Ben-Chang Shia[9,10*], PhD; Linglong Ye[11*], PhD

[1]School of Statistics, University of International Business and Economics, Beijing, China
[2]Guanghua School of Management, Peking University, Beijing, China
[3]College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan
[4]Taipei Cancer Center, Taipei Medical University, Taipei, Taiwan
[5]Department of Pediatrics, City of Hope National Medical Center, Duarte, CA, United States
[6]Department of Pediatrics, Taipei Medical University Hospital, Taipei, Taiwan
[7]Department of Pediatrics, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan
[8]International Ph.D. Program in Biotech and Healthcare Management, College of Management, Taipei Medical University, Taipei, Taiwan
[9]Graduate Institute of Business Administration, College of Management, Fu Jen Catholic University, New Taipei City, Taiwan
[10]Artificial Intelligence Development Center, Fu Jen Catholic University, New Taipei City, Taiwan
[11]School of Public Affairs, Xiamen University, Xiamen, China
*these authors contributed equally

Corresponding Author:
Linglong Ye, PhD
School of Public Affairs
Xiamen University
No.422, Siming South Road
Xiamen, 361005
China
Phone: 86 5922182783
Fax: 86 5922183191
Email: leyloria@gmail.com

Abstract

Background: The use of social media in communications regarding cancer prevention is rapidly growing. However, less is known about the general population's social media use related to cancer screening awareness and behavior for different cancers.

Objective: We aimed to examine the relationship between social media use and cancer screening awareness and behavior among people without a cancer diagnosis.

Methods: Data were collected from the Health Information National Trends Survey 5 Cycle 1 to 3 in the United States (n=12,227). Our study included 10,124 participants without a cancer diagnosis and 3 measures of screening awareness (those who had heard of hepatitis C virus [HCV], human papillomavirus [HPV], and the HPV vaccine) and 4 measures of behavior (those who had prostate-specific antigen tests, Papanicolaou tests for cervical cancer, as well as breast cancer and colon cancer tests). Propensity-score matching was conducted to adjust for the sociodemographic variables between the social media user and nonuser participants. Multivariable logistic regression was used to assess the association of social media use by gender. Jackknife replicate weights were incorporated into the analyses.

Results: Of the 3794 matched participants, 1861 (57.6% weighted) were male, and the mean age was 55.5 (SD 0.42) years. Compared to social media nonusers, users were more likely to have heard of HCV (adjusted odds ratio [aOR]=2.27, 95% CI, 1.29-3.98 and aOR=2.86, 95% CI, 1.51-5.40, for male and female users, respectively) and HPV (aOR=1.82, 95% CI, 1.29-2.58 and aOR=2.35, 95% CI, 1.65-3.33, for male and female users, respectively). In addition, female users were more likely to have heard of the HPV vaccine (aOR=2.06, 95% CI, 1.41-3.00). No significant associations were found between social media use and prostate-specific antigen tests in males, Papanicolaou tests and breast cancer tests in females, or colon cancer tests in both male and female users.

Conclusions: While social media services can potentially promote cancer screening awareness in the general population, but they did not improve screening behavior after adjusting for socioeconomic status. These findings strengthened our understanding of social media use in targeting health communications for different cancers.

(J Med Internet Res 2021;23(8):e26395) doi:10.2196/26395

KEYWORDS

social media; cancer screening awareness; cancer screening behavior; gender-specific effects; propensity-score matching; general population

Introduction

The introduction of screening for early prostate, colorectal, breast, and cervical cancer detection has significantly reduced mortality rates over the past few decades [1]. However, cancer remains the second leading cause of death in the United States [2], and premature cancer deaths resulted in US $94.4 billion lost earnings among people aged 16 to 84 years in 2015 [3]. Therefore, there is growing concern regarding the low proportion of cancer screening awareness and behavior in the general population and the relative disparities in cancer screening awareness and behavior due to race and ethnicity and socioeconomic status.

A recent National Health Interview Survey reported that the utilization of recommended cancer screenings is far lower than the Healthy People 2020 targets for the nation [4]. However, cancer screening awareness and behavior disparities exist according to cancer types, race and ethnicity, socioeconomic status, and other health care access factors [1,4,5]. For example, from 2005 to 2015, colorectal cancer screenings have increased steadily, but prostate, breast, and cervical cancer screenings have declined [6]. In addition, prostate, cervical, breast, and colon cancer screening rates are lower among Hispanic and Asian groups than non-Hispanic White and Black groups, declining with decreasing education levels [6]. Therefore, new channels for informing target populations are needed to improve the cancer screening awareness and behavior regarding different cancers and reduce the relative disparities.

The use of social media has increased over the past decade and has become a new channel for promoting cancer prevention [7]. Compared to older approaches to cancer interventions with unidirectional and paternalistic outreach barriers, contemporary social media is easy to access, bidirectional, interactive, and patient-driven [7,8]. Social media services are widely used, varied, and continually innovating, offering substantial opportunities for health communication [9]. Given these inherent advantages of social media, there has been rapid growth in the use of different social media services by health care providers and the general population to communicate information regarding cancer prevention, such as basic cancer knowledge, healthy lifestyles, and the importance of cancer screenings.

Previous research on the use of social media in cancer prevention has primarily focused on educational material [7]. In most previous studies, general information regarding cancer screening has been reported as a key component of cancer prevention education. However, a single study has addressed the detailed descriptions of bowel preparation for colorectal cancer screening [7,10]. Among these studies, cancer screening awareness and behavior for different cancers were not thoroughly examined [7]. Thus, more insight is needed to understand better the effects of social media on peoples' awareness and behavior concerning the screening for different cancers.

Despite the diversity of social media services, the previous studies mainly used specific types of social media, including Facebook, Twitter, and WeChat (a widely used social media app in China) [10-12]. Although one qualitative study reported that Facebook use helped promote breast cancer screening awareness [12], inconsistent effects of different social media services were found for the screening behavior for different cancer types [10,11]. Rosemary et al [11] showed that using Twitter did not improve specific behaviors regarding breast cancer prevention, including screening. In contrast, one study revealed an improvement in colorectal screening behavior through the use of WeChat [10]. However, the discussions are still too limited to provide comprehensive information to determine and compare the effects of social media on cancer screening awareness and behavior for different cancer types.

The variety of evolving social media services makes it necessary to examine how the general use of social media impacts cancer screening awareness and behavior for different cancer types. Moreover, most users of different social media services vary across racial and ethnic minority and socioeconomic groups. Only certain populations were included in previous prevention studies using specific types of social media [7,10-12]. Even less is known about the general population's social media use related to cancer screening awareness and behavior for different types of cancer.

This study aimed to assess how the general use of social media impacted people's cancer screening awareness and behavior for different cancer types in a nationally representative sample of US adults without a cancer diagnosis. In addition, because of the inherent heterogeneity of different cancer types between males and females, the gender-specific effects of social media use were examined. By evaluating social media use in the general population, this study furthers our understanding of the relationship between health communications using social media and health-promoting behaviors.

Methods

Data Sources

This study adopted data from the National Cancer Institute (NCI) Health Information National Trends Survey (HINTS), including HINTS 5 Cycle 1 (2017, n=3285), HINTS 5 Cycle 2 (2018, n=3504), and HINTS 5 Cycle 3 (2019, n=5438). The HINTS is a nationwide survey of noninstitutionalized individuals aged 18 years and older in the United States and has been conducted every 1 to 2 years by the NCI since 2003. It uses a probability-based sampling with a 2-level stratified design, considering areas with high or low concentrations of minorities. Data from the 3 surveys were collected by mail or Web Pilot from January to May 2017, 2018, and 2019, respectively. Data were available in a publicly accessible repository that does not issue DOIs. Publicly available data sets were analyzed in this study. More detailed information on the study design and high-quality HINTS data have been published elsewhere [13]. After excluding those individuals with a cancer diagnosis, 10,124 participants remained. Participants with missing data for any variables were excluded from the final analyses, resulting in a sample of 7090 social media users and 2775 social media nonusers.

Measurements

Social Media Use

Social media use was assessed using all 5 subquestions of the B14 question in the HINTS questionnaire (“Sometimes people use the internet to connect with other people online through social networks like Facebook or Twitter. This is often called ‘social media.’”). The 5 subquestions asked, “In the past 12 months, have you used the internet for any of the following reasons: (1) to visit a social networking site, such as Facebook or LinkedIn; (2) to share health information on social networking sites, such as Facebook or Twitter; (3) to write in an online diary or blog (ie, weblog); (4) to participate in an online forum or support group for people with similar health or medical issues; or (5) to watch a health-related video on YouTube?” The participants answered either “yes” or “no,” and those using social media were defined by a “yes” response to any of the 5 questions.

Cancer Screening Awareness and Behavior

The measures of cancer screening awareness consisted of 3 dependent variables, which were assessed by 3 questions: (1) “Have you ever heard of the hepatitis C virus (also known as Hep C or HCV)?”; (2) “Have you ever heard of HPV?”; (3) “Before today, have you ever heard of the cervical cancer vaccine or HPV shot?” Participants indicated their response with either a “yes” (1) or “no” (0).

Cancer screening behavior measures contained 4 dependent variables defined by 4 questions: (1) “Have you ever had a prostate-specific antigen (PSA) test?” (1=yes; 0=no); (2) “How long ago did you have your most recent Papanicolaou (Pap) test to check for cervical cancer?” (1=I have had a Pap test [a year ago or less, 1 to 2 years ago, 2 to 3 years ago, 3 to 5 years ago, or more than 5 years ago]; 0=I have never had a Pap test); (3) “When did you have your most recent mammogram to check for breast cancer, if ever?” (1=I have had a mammogram [a year ago or less, 1 to 2 years ago, 2 to 3 years ago, 3 to 5 years ago, or more than 5 years ago]; 0=I have never had a mammogram); (4) “Have you ever had one of these tests, including colonoscopy, sigmoidoscopy, and stool blood test to check for colon cancer?” (1=yes; 0=no).

Due to the differences in the survey content between the different Cycles, the data for HCV were from Cycle 2 (2018) and Cycle 3 (2019), the data for HPV and the HPV vaccine were from Cycles 1 to 3 (2017-2019), the data for PSA testing were from Cycle 1 (2017) and Cycle 3 (2019), the data for Pap and breast cancer testing were from Cycles 1 to 3 (2017-2019), and the data for colon cancer testing were from Cycle 2 (2018) and Cycle 3 (2019).

Sociodemographic Covariates

The sociodemographic characteristics in this study included self-reported measures of gender (male or female), age, race and ethnicity (non-Hispanic White or racial and ethnic minority), education level (high school or less or more than high school), income (less than US $20,000 or US $20,000 or more annually), and geographic area (nonmetropolitan or metropolitan). Measurements of the geographic area were obtained by 9 metropolitan codes corresponding to the Rural-Urban Continuum Codes with a range of 1 to 9 [14]. Based on the cutoff points adopted by the United States Department of Agriculture and previous HINTS studies, this study divided the geographic areas into metropolitan (codes 1 to 3) and nonmetropolitan (codes 4 to 9) areas [14-16].

Statistical Analysis

This study compared the sociodemographic characteristics of social media users and nonusers via chi-squared tests for categorical variables and 2-tailed *t*-tests for continuous variables. Propensity-score matched analysis was carried out to adjust the sociodemographic variables between the social media users and nonusers. Confounders used for matching social media users and nonusers included gender, age, race and ethnicity, education, income, and geographic area. Given the gender differences for different cancer types, the differences in cancer screening awareness and behavior were compared by the social media use of males and females. Crude odds ratios (cOR), adjusted odds ratios (aOR), and their 95% CIs stratified by gender were computed using univariate and multivariable logistic regression analyses to clarify the impact of social media use on each dependent variable. This study adjusted the multivariable analyses for potential confounders, including gender, age, race and ethnicity, education, income, and geographic area.

All statistical analyses were carried out using SAS (version 9.4; SAS Institute). Given the complex survey design of the HINTS, jackknife replicate weights were incorporated into the analyses to obtain population-level estimates. The jackknife method creates a set of replicate samples from the original sample and provides an estimate of the variable and its variance of interest [17,18]. A $P<.05$ was considered statistically significant.

Results

Sociodemographic Characteristics and Social Media Use

All percentages, means, and standard errors reported in this section are the weighted values. Among 10,124 participants without a diagnosis of cancer, 50% (n=4103) were male, and the average age was 48 (SD 0.20) years (Table 1). Nearly two-thirds (n=5551, 59%) of the participants were non-Hispanic White, and about one-third (n=2496, 31%) had a high school education or less. The income of about 18% (n=1715) of the participants was less than US $20,000 annually. In addition, 14% (n=1246) of the population lived in nonmetropolitan areas. Among the 9868 participants with complete information, 72% (n=7093) used social media. There were statistically significant differences in all sociodemographic characteristics between the social media user versus nonuser participants (all *P*<.001). Compared to social media nonusers, the social media users were more likely to be female (56%, n=1291), tended to be younger (mean age 44 years vs 59 years), non-Hispanic White (n=4129, 61% vs n=1344, 54%), were educated above a high school level (76%, n=5765 vs n=1580, 50%), and were more likely to live in metropolitan areas (n=6298, 88% vs n=2361, 83%).

Table 1. Sociodemographic characteristics for social media users and nonusers. All percentages, means, and standard errors reported in the table are the weighted values.

Characteristics	Total (N=10,124)	Social media use (n=9868)		
		Use (n=7093)	Nonuse (n=2775)	*P* value[a]
Gender, n (%)				
Male	4103 (49.4)	2723 (47.3)	1291 (56.4)	<.001
Female	5895 (50.6)	4308 (52.7)	1433 (43.6)	
Age (years), mean (SE)	47.6 (0.20)	44.1 (0.25)	58.8 (0.56)	<.001
Race and ethnicity, n (%)				
Non-Hispanic White	5551 (58.6)	4129 (60.7)	1344 (53.5)	<.001
Racial and ethnic minority	4573 (41.4)	2964 (39.3)	1431 (46.5)	
Education, n (%)				
≤High school	2496 (30.7)	1277 (24.4)	1108 (49.6)	<.001
>High school	7479 (69.3)	5765 (75.6)	1580 (50.4)	
Income (US), n (%)				
<$20,000	1715 (18.0)	909 (14.9)	732 (27.6)	<.001
≥$20,000	7397 (82.0)	5662 (85.1)	1616 (72.4)	
Geographic area, n (%)				
Nonmetropolitan	1246 (13.5)	795 (12.4)	414 (17.5)	<.001
Metropolitan	8878 (86.5)	6298 (87.6)	2361 (82.5)	

[a] Chi-squared tests for categorical variables and *t*-tests for continuous variables.

Sociodemographics of Participants After Propensity-Score Matching

After propensity-score matching, a total of 3794 (33%) participants (1897 social media users and nonusers each) remained (Table 2). Of those 1864 (58%) were male, 2110 (59%) were non-Hispanic white, 1278 (45%) had a high school education or less, 907 (24%) had an annual income below US $20,000, and 528 (16%) lived in nonmetropolitan areas. No statistically significant differences between social media users and nonusers were found in any of the sociodemographic characteristics.

Table 2. Sociodemographic characteristics between social media users and nonusers matched by the propensity-score method. All percentages, means, and standard errors reported in the table are the weighted values.

Characteristics	Total (n=3794)	Social media use (n=3794)		
		Use (n=1897)	Nonuse (n=1897)	*P* value[a]
Gender, n (%)				
Male	1861 (57.6)	941 (57.2)	920 (58.0)	.77
Female	1933 (42.4)	956 (42.8)	977 (42.0)	
Age (years), mean (SE)	55.5 (0.42)	55.6 (0.49)	55.5 (0.69)	.98
Race/ethnicity, n (%)				
Non-Hispanic White	2110 (59.1)	1066 (59.2)	1044 (59.0)	.93
Racial and ethnic minority	1684 (40.9)	831 (40.8)	853 (41.0)	
Education, n (%)				
≤High school	1278 (45.3)	656 (46.8)	622 (43.9)	.33
>High school	2516 (54.7)	1241 (53.2)	1275 (56.1)	
Income (US), n (%)				
<$20,000	907 (23.5)	442 (23.9)	465 (23.0)	.74
≥$20,000	2887 (76.5)	1455 (76.1)	1432 (77.0)	
Geographic area, n (%)				
Nonmetropolitan	528 (16.4)	287 (17.2)	241 (15.6)	.42
Metropolitan	3266 (83.6)	1610 (82.8)	1656 (84.4)	

[a] Chi-squared tests for categorical variables and t-tests for continuous variables.

Cancer Screening Awareness and Behavior by Social Media Use and Gender

Male social media users, compared to nonusers, were more likely to have heard of HCV (n=394, 92% vs n=321, 84%; *P*=.02) and HPV (n=502, 56% vs n=398, 43%; *P*=.003). However, awareness of the HPV vaccine was not significantly different between male social media users and nonusers (n=420, 48% vs n=369, 41%, respectively; *P*=.12). For all 1861 male participants, including both social media users and nonusers, about 815 (49%) had taken PSA tests (*P*=.97), and 488 (55%) of the social media users had taken colon cancer tests compared to 433 (47%) social media nonusers (*P*=.09; Table 3).

For the 1933 female participants, the social media users, compared to nonusers, were more likely to have heard of HCV (89%, n=402 vs 75%, n=336; *P*<.001), and HPV (72%, n=683 vs 56%, n=566; *P*≤.001), and HPV vaccine (72%, n=664 vs 58%, n=577; *P*≤.001). In addition, 97% (n=913) social media users and 94% (n=921) nonusers had taken Pap tests (*P*=.31), 87% (n=849) social media users and 85% (n=874) nonusers had taken breast cancer tests (*P*=.43), and 66% (n=514) social media users and 64% (n=479) nonusers had taken colon cancer tests (*P*=.61).

Table 3. Differences in cancer screening awareness and behavior by social media use and gender. All percentages, means, and standard errors reported in the table are the weighted values.

Characteristics	Social media use					
	Male (n=1861)			Female (n=1933)		
	Use (n=941)	Nonuse (n=920)	*P* value[a]	Use (n=956)	Nonuse (n=977)	*P* value
Awareness						
Heard of HCV[b] (yes), n (%)	394 (91.7)	321 (83.5)	.02	402 (89.3)	336 (75.0)	.002
Heard of HPV[c] (yes), n (%)	502 (56.4)	398 (43.4)	.003	683 (72.3)	566 (55.5)	<.001
Heard of HPV vaccine (yes), n (%)	420 (48.0)	369 (40.5)	.12	664 (71.7)	577 (57.7)	<.001
Behavior						
Had PSA[d] test (yes), n (%)	427 (49.0)	388 (49.2)	.97	N/A	N/A	
Had Pap[e] test (yes), n (%)	N/A	N/A		913 (96.5)	921 (94.1)	.31
Had breast cancer test (yes), n (%)	N/A	N/A		849 (87.1)	874 (85.0)	.43
Had colon cancer test (yes), n (%)	488 (54.8)	443 (46.7)	.09	514 (65.7)	479 (64.0)	.61

[a] Chi-squared tests for categorical variables and t-tests for continuous variables.

[b] HCV: hepatitis C virus.

[c] HPV: human papillomavirus.

[d] PSA: prostate-specific antigen.

[e] Pap: Papanicolaou.

Impact of Social Media Use on Awareness and Behavior of Cancer Screening

Table 4 reports the relationship between social media use and cancer screening awareness and behavior in male and female participants presented by weighted analyses. Male social media users were more likely to have heard of HCV (cOR 2.17, 95% CI 1.25-3.77; aOR 2.27, 95% CI 1.29-3.98) and HPV (cOR 1.68, 95% CI 1.20-2.36; aOR 1.82, 95% CI 1.29-2.58) compared to nonusers. No statistically significant effects of social media use on awareness of the HPV vaccine and taking PSA tests or colon cancer tests were observed in males.

Table 4. Univariate and multivariable logistic regression assessing the impact of social media use on cancer screening awareness and behavior. Weighted analyses are presented.

Characteristics	Male (n=1861)				Female (n=1933)			
	cOR[a] (95% CI)	*P* value	aOR[b] (95% CI)	*P* value	cOR (95% CI)	*P* value	aOR (95%CI)	*P* value
Awareness								
Heard of HCV[c]	2.17 (1.25-3.77)	.007	2.27 (1.29-3.98)	.005	2.80 (1.57-4.99)	<.001	2.86 (1.51-5.40)	.002
Heard of HPV[d]	1.68 (1.20-2.36)	.003	1.82 (1.29-2.58)	.001	2.10 (1.53-2.87)	<.001	2.35 (1.65-3.33)	<.001
Heard of HPV vaccine	1.35 (0.92-1.98)	.12	1.42 (0.98-2.06)	.07	1.85 (1.34-2.57)	<.001	2.06 (1.41-3.00)	<.001
Behavior								
Had PSA[e] test	0.99 (0.68-1.46)	.97	1.17 (0.73-1.88)	.52	N/A		N/A	
Had Pap[f] Test	N/A		N/A		1.72 (0.57-5.19)	.33	1.54 (0.58-4.10)	.38
Had breast cancer test	N/A		N/A		1.20 (0.76-1.88)	.43	1.06 (0.70-1.60)	.78
Had colon cancer test	1.38 (0.95-2.02)	.09	1.58 (0.98-2.54)	.06	1.08 (0.80-1.45)	.62	1.24 (0.88-1.74)	.21

[a] cOR: crude odds ratio.

[b] aOR: adjusted odds ratio.

[c] HCV: hepatitis C virus.

[d] HPV: human papillomavirus.

[e] PSA: prostate-specific antigen.

[f] Pap: Papanicolaou.

Female social media users were more than twice as likely to have heard of HCV (cOR 2.80, 95% CI 1.57-4.99; aOR 2.86, 95% CI 1.51-5.40), HPV (cOR 2.10, 95% CI 1.53-2.87; aOR 2.35, 95% CI 1.65-3.33) or the HPV vaccine (cOR 1.85, 95% CI 1.34-2.57; aOR 2.06, 95% CI 1.41-3.00) compared to nonusers. There was no statistically significant impact of social media use on taking Pap tests, breast cancer tests, and colon cancer tests.

Discussion

Principal Results

Our study conducted a comprehensive evaluation of the effects of general social media use on cancer screening awareness and behavior among US adults without a cancer diagnosis. Although the impact of social media on screening awareness varied across different cancer types and genders, we showed social media could promote cancer screening awareness in the general population. However, it did not improve screening behavior after adjusting for race and ethnicity and socioeconomic status. In light of the growing use of social media by health care providers and the general population, the findings of this study can contribute to targeted messaging for cancer prevention and reduce disparities between the different groups.

After matching and controlling for sociodemographic characteristics, we found that the general social media use had a significant impact on the awareness of HPV in both male and female adults. In contrast, its effect on HPV vaccine awareness was only significant in female adults. Similar to an online cervical cancer prevention study, HPV awareness can be increased through social media [19]. Persistent infection with specific HPV subtypes accounts for almost all cervical cancers in women, practically all anal cancers (over 9 out of 10), most vaginal, oropharyngeal, vulvar, and penile cancers (between two-thirds and three-quarters), and some oral cavity and laryngeal cancers in both men and women [20]. In addition to cervical cancer, most HPV-associated cancers could be reduced by current vaccines [20]. Hence, HPV vaccination is recommended for both males and females at specific ages. The American Cancer Society suggests that all children aged 11 or 12 years should be vaccinated against HPV infections. HPV vaccination with Gardasil 9 (Merck) has been approved for men and women aged 26 to 45 years by the United States Food and Drug Administration since 2018 [1,21]. An Australian study suggested that Twitter-derived measures of information exposure were associated with HPV vaccine coverage for both adolescent males and females [22]. Hence, in addition to communicating about HPV vaccination and cervical cancer prevention, social media can provide additional health information about other HPV-associated cancers and target the male population.

A significant impact of social media use on HCV awareness was also found for both male and female adults in this study. A southeast Michigan study on health disparities in hepatitis C revealed a suboptimal impact of multiple sociodemographic factors on hepatitis C screening and care [23]. Through the internet and a personal health record training program, health-related internet use increased among low-income patients with HCV, and their self-efficacy and patient activation were also improved [24]. As a patient-driven internet service, social media emerges as a tool linking the general population with HCV awareness. However, its impact on improving hepatitis C screening and care among specific groups at high risk of HCV requires further research.

Social media use did not significantly impact cancer screening behavior, including taking PSA tests in male adults, Pap tests and breast cancer tests in female adults, or colon cancer tests in both male and female adults. In contrast, previous studies regarding specific social media services in the United States or China showed inconsistent effects of social media on cancer screening behavior [10-12]. These results may be due partly to misinformation in cancer communications commonly shared via social media.

There are increasing concerns about misleading or inaccurate medical science information on social media [25-27]. For example, a recent study on the accuracy of genitourinary malignancy articles shared on social media showed inaccurate and misleading articles were shared at a significantly higher rate on average than accurate articles, and 1 in 10 articles about PSA testing contained misinformation [28].

In addition, a lack of consensus on the clinical use of cancer screening and treatment may cause unnecessary panic when shared via social media. The so-called gray area of PSA testing with a high rate of false positives has resulted in its questionable clinical usefulness in Italy [29]. Prostate cancer tends to develop slowly, and its overdiagnosis and overtreatment are controversial [30]. Overdiagnosis and overtreatment have the potential for unnecessary side effects, including physical and psychological harms. The decrements in quality of life from side effects may offset some of the gains in length of life obtained from screening [31,32].

Given that social media amplifies scientific information and misinformation [25], a Japanese study [33] has revealed the information war for and against cancer screening messages has begun online. Therefore, accurate information with more engagement from scientists or professional institutions on social media is urgently needed to prevent misinformation about cancer screening from expanding and to achieve the screening target rates recommended in Healthy People 2030. In addition, collaboration with celebrities diagnosed with cancer may trigger substantial social media interest compared to traditional efforts for raising cancer screening awareness and affect screen behaviors, allowing health care providers to leverage social media methods like celebrities.

In terms of routine cancer care, PSA or colon cancer tests for males and Pap tests and breast cancer tests for females are part of routine check-ups. However, it may not be easy to change the behavior of people with negative perceptions regarding cancer screenings. As several HPV vaccine studies have pointed out, prior exposure to negative information was correlated with the later expression of negative opinions [34], and among Twitter users in Canada, the United Kingdom, and Australia, those with negative or opposing views about vaccinations were more likely to be better connected than those with positive views [35], leading to more barriers to improving routine cancer screening. As for cancers and tests that are not part of routine care, such as skin cancer [31], social media could be used to increase prior exposure to positive views for targeted populations to raise screening rates.

The findings in this study also suggest that among social media users, increased health information awareness did not lead to increased cancer screening behavior. The gap between “knowing” and “doing” in cancer screening behavior may be associated with the social characteristics of social media. Vaterlaus et al [36] reported a significant impact of social media on diet and exercise, which are commonly shared on social media. Compared to diet and exercise, health communication about cancer screening targets relatively small groups and has less engagement. Breast Cancer Awareness Month and associated events on Twitter revealed that a strategic communication plan calls for ongoing social media conversations on breast cancer and screening mammograms to help increase breast cancer screening rates [11].

The reach of social media is substantial, and the demographics of health-related users vary across different social media platforms [37]. It is valuable to promote cancer screening through various social media platforms according to the demographics of their health-related users to increase cancer screening rates for certain socioeconomic groups and hard-to-reach populations. Combined with ads targeted toward specific audiences, social media platforms could provide potentially innovative and effective ways to communicate positive health information.

Limitations

This study had several limitations. First, this study was based on the general use of social media rather than health-related use, which may result in measurement errors. Given the wide use of social media, it is difficult to separate health communications from the general use of social media. For instance, COVID-19 dominated social media communications at the beginning of 2020. Furthermore, the degree of social network use (eg, frequency and duration) was not quantified, and thus, the time-dependent effects of social media use could not be measured. Second, this study defined the use of Pap tests and breast cancer tests via “yes” or “no” responses among all the participants, without considering the recommended screening frequency by age. Moreover, the patterns of social media use vary across different age groups. Further studies on the effects of social media use on the frequency of cancer screening awareness and behavior at different ages should be conducted.

Conclusions

This study identified the association between social media use and cancer screening awareness and behavior among people without a cancer diagnosis after matching and controlling for sociodemographic factors. Social media use was associated with HPV and HCV awareness among both male and female participants, and the awareness of the HPV vaccine among female participants. No significant correlation was observed between social media use and screening behavior, including PSA, Pap, breast cancer, or colon cancer test-taking. This study provided a comprehensive assessment of the association between general social media use and cancer screening awareness and behavior for different types of cancer in the general population. Our study suggests that health communication using social media can effectively impact the awareness of specific cancers. Still, a more innovative and targeted approach is needed, including accurately delivering messages to hard-to-reach populations and improving specific screening behaviors.

Acknowledgments

This work was supported by the University of International Business and Economics (UIBE) Huiyuan outstanding young scholars research fund (grant 17YQ15) and the Fundamental Research Funds for the Central Universities in UIBE (grant CXTD10-10).

Conflicts of Interest

None declared.

References

1. Smith R, Andrews K, Brooks D, Fedewa S, Manassaram‐Baptiste D, Saslow D, et al. Cancer screening in the United States, 2019: A review of current American Cancer Society guidelines and current issues in cancer screening. CA A Cancer J Clin 2019 Mar 15;69(3):184-210. [doi: 10.3322/caac.21557]
2. Siegel R, Miller KD, Jemal A. Cancer statistics, 2019. CA A Cancer J Clin 2019 Jan 08;69(1):7-34. [doi: 10.3322/caac.21551]
3. Islami F, Miller KD, Siegel RL, Zheng Z, Zhao J, Han X, et al. National and State Estimates of Lost Earnings From Cancer Deaths in the United States. JAMA Oncol 2019 Sep 12;5(9):e191460. [doi: 10.1001/jamaoncol.2019.1460]
4. White A, Thompson TD, White MC, Sabatino SA, de Moor J, Doria-Rose PV, et al. Cancer Screening Test Use - United States, 2015. MMWR Morb Mortal Wkly Rep 2017 Mar 03;66(8):201-206 [FREE Full text] [doi: 10.15585/mmwr.mm6608a1] [Medline: 28253225]
5. Goel MS, Wee CC, McCarthy EP, Davis RB, Ngo-Metzger Q, Phillips RS. Racial and ethnic disparities in cancer screening the importance of foreign birth as a barrier to care. J Gen Intern Med 2003 Dec;18(12):1028-1035. [doi: 10.1111/j.1525-1497.2003.20807.x]
6. Wender RC, Brawley OW, Fedewa SA, Gansler T, Smith RA. A blueprint for cancer screening and early detection: Advancing screening’s contribution to cancer control. CA A Cancer J Clin 2018 Nov 19;69(1):50-79. [doi: 10.3322/caac.21550]

7. Han CJ, Lee YJ, Demiris G. Interventions Using Social Media for Cancer Prevention and Management. Cancer Nurs 2018;41(6):E19-E31. [doi: 10.1097/ncc.0000000000000534]
8. Attai DJ, Cowher MS, Al-Hamadani M, Schoger JM, Staley AC, Landercasper J. Twitter Social Media is an Effective Tool for Breast Cancer Patient Education and Support: Patient-Reported Outcomes by Survey. J Med Internet Res 2015 Jul 30;17(7):e188. [doi: 10.2196/jmir.4721]
9. Huo J, Desai R, Hong Y, Turner K, Mainous AG, Bian J. Use of Social Media in Health Communication: Findings From the Health Information National Trends Survey 2013, 2014, and 2017. Cancer Control 2019 Apr 18;26(1):107327481984144. [doi: 10.1177/1073274819841442]
10. Kang X, Zhao L, Leung F, Luo H, Wang L, Wu J, et al. Delivery of Instructions via Mobile Social Media App Increases Quality of Bowel Preparation. Clinical Gastroenterology and Hepatology 2016 Mar;14(3):429-435.e3. [doi: 10.1016/j.cgh.2015.09.038]
11. Thackeray R, Burton SH, Giraud-Carrier C, Rollins S, Draper CR. Using Twitter for breast cancer prevention: an analysis of breast cancer awareness month. BMC Cancer 2013 Oct 29;13(1). [doi: 10.1186/1471-2407-13-508]
12. Abramson K, Keefe B, Chou WS. Communicating About Cancer Through Facebook: A Qualitative Analysis of a Breast Cancer Awareness Page. Journal of Health Communication 2014 Dec 13;20(2):237-243. [doi: 10.1080/10810730.2014.927034]
13. Finney RL, Blake K, Skolnick V, Davis T, Moser R, Hesse B. Data resource profile: the national Cancer Institute’s Health Information National Trends Survey (HINTS). International journal of epidemiology 2019:17-17j. [doi: 10.1093/ije/dyz083]
14. United SDOA. Rural-urban continuum codes: Documentation. Rural-Urban Continuum Codes. 2013. URL: https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/documentation.aspx [accessed 2020-04-25]
15. Vereen RN, Westmaas JL, Bontemps-Jones J, Jackson K, Alcaraz KI. Trust of Information about Tobacco and E-Cigarettes from Health Professionals versus Tobacco or Electronic Cigarette Companies: Differences by Subgroups and Implications for Tobacco Messaging. Health Communication 2018 Nov 13;35(1):89-95. [doi: 10.1080/10410236.2018.1544875]
16. Heffner JL, Mull KE. Smartphone Ownership Among US Adult Cigarette Smokers: 2014 Health Information National Trends Survey (HINTS) Data. J Med Internet Res 2017 Aug 31;19(8):e305. [doi: 10.2196/jmir.7953]
17. Chandrasekaran R, Katthula V, Moustakas E. Patterns of Use and Key Predictors for the Use of Wearable Health Care Devices by US Adults: Insights from a National Survey. J Med Internet Res 2020 Oct 16;22(10):e22443. [doi: 10.2196/22443]
18. Lavrakas PJ. Jackknife variance estimation. In: Encyclopedia of Survey Research Methods. Newbury Park, California: Sage publications; 2008. [doi: 10.4135/9781412963947.n257]
19. Lyson HC, Le GM, Zhang J, Rivadeneira N, Lyles C, Radcliffe K, et al. Social Media as a Tool to Promote Health Awareness: Results from an Online Cervical Cancer Prevention Study. J Canc Educ 2018 Jun 11;34(4):819-822. [doi: 10.1007/s13187-018-1379-8]
20. Saraiya M, Unger E, Thompson T, Lynch C, Hernandez B, Lyu C. US assessment of HPV types in cancers: implications for current and 9-valent HPV vaccines. JNCI: Journal of the National Cancer Institute 2015;107(6):djv086. [doi: 10.1093/jnci/djv086]
21. Saslow D, Andrews KS, Manassaram-Baptiste D, Loomer L, Lam KE, Fisher-Borne M, et al. Human papillomavirus vaccination guideline update: American Cancer Society guideline endorsement. CA: A Cancer Journal for Clinicians 2016 Jul 19;66(5):375-385. [doi: 10.3322/caac.21355]
22. Dyda A, Shah Z, Surian D, Martin P, Coiera E, Dey A, et al. HPV vaccine coverage in Australia and associations with HPV vaccine information exposure among Australian Twitter users. Human Vaccines & Immunotherapeutics 2019 Apr 12;15(7-8):1488-1495. [doi: 10.1080/21645515.2019.1596712]
23. Bourgi K, Brar I, Baker-Genaw K. Health Disparities in Hepatitis C Screening and Linkage to Care at an Integrated Health System in Southeast Michigan. PLoS ONE 2016 Aug 15;11(8):e0161241. [doi: 10.1371/journal.pone.0161241]
24. McInnes D, Solomon J, Shimada S, Petrakis B, Bokhour B, Asch S. Developmentevaluation of an internetpersonal health record training program for low-income patients with HIV or hepatitis C. Medical care 2013:S62-S66. [doi: 10.1097/mlr.0b013e31827808bf]
25. Merchant RM, Asch DA. Protecting the Value of Medical Science in the Age of Social Media and “Fake News”. JAMA 2018 Dec 18;320(23):2415-2416. [doi: 10.1001/jama.2018.18416]
26. 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]
27. Wang Y, McKee M, Torbica A, Stuckler D. Systematic Literature Review on the Spread of Health-related Misinformation on Social Media. Social Science & Medicine 2019 Nov;240:112552. [doi: 10.1016/j.socscimed.2019.112552]
28. Alsyouf M, Stokes P, Hur D, Amasyali A, Ruckle H, Hu B. ‘Fake News’ in urology: evaluating the accuracy of articles shared on social media in genitourinary malignancies. BJU Int 2019 May 13;124(4):701-706. [doi: 10.1111/bju.14787]
29. Perdonà S, Cavadas V, Di Lorenzo G, Damiano R, Chiappetta G, Del Prete P, et al. Prostate Cancer Detection in the “Grey Area” of Prostate-Specific Antigen Below 10 ng/ml: Head-to-Head Comparison of the Updated PCPT Calculator and Chun’s Nomogram, Two Risk Estimators Incorporating Prostate Cancer Antigen 3. European Urology 2011 Jan;59(1):81-87. [doi: 10.1016/j.eururo.2010.09.036]

30. Loeb S, Bjurlin MA, Nicholson J, Tammela TL, Penson DF, Carter HB, et al. Overdiagnosis and Overtreatment of Prostate Cancer. European Urology 2014 Jun;65(6):1046-1055. [doi: 10.1016/j.eururo.2013.12.062]
31. Falzone AE, Brindis CD, Chren M, Junn A, Pagoto S, Wehner M, et al. Teens, Tweets, and Tanning Beds: Rethinking the Use of Social Media for Skin Cancer Prevention. American Journal of Preventive Medicine 2017 Sep;53(3):S86-S94. [doi: 10.1016/j.amepre.2017.04.027]
32. Heijnsdijk EA, Wever EM, Auvinen A, Hugosson J, Ciatto S, Nelen V, et al. Quality-of-Life Effects of Prostate-Specific Antigen Screening. N Engl J Med 2012 Aug 16;367(7):595-605. [doi: 10.1056/nejmoa1201637]
33. Okuhara T, Ishikawa H, Okada M, Kato M, Kiuchi T. Assertions of Japanese Websites for and Against Cancer Screening: a Text Mining Analysis. Asian Pac J Cancer Prev 2017 Apr 01;18(4):1069-1075 [FREE Full text] [doi: 10.22034/APJCP.2017.18.4.1069] [Medline: 28547943]
34. Dunn AG, Leask J, Zhou X, Mandl KD, Coiera E. Associations Between Exposure to and Expression of Negative Opinions About Human Papillomavirus Vaccines on Social Media: An Observational Study. J Med Internet Res 2015 Jun 10;17(6):e144. [doi: 10.2196/jmir.4343]
35. Shapiro GK, Surian D, Dunn AG, Perry R, Kelaher M. Comparing human papillomavirus vaccine concerns on Twitter: a cross-sectional study of users in Australia, Canada and the UK. BMJ Open 2017 Oct 05;7(10):e016869 [FREE Full text] [doi: 10.1136/bmjopen-2017-016869] [Medline: 28982821]
36. Vaterlaus JM, Patten EV, Roche C, Young JA. #Gettinghealthy: The perceived influence of social media on young adult health behaviors. Computers in Human Behavior 2015 Apr;45:151-157. [doi: 10.1016/j.chb.2014.12.013]
37. Sadah SA, Shahbazi M, Wiley MT, Hristidis V. A Study of the Demographics of Web-Based Health-Related Social Media Users. J Med Internet Res 2015 Aug 06;17(8):e194. [doi: 10.2196/jmir.4308]

Abbreviations

aOR: adjusted odds ratios
cOR: crude odds ratios
HCV: hepatitis C virus
HINTS: Health Information National Trends Survey
HPV: human papillomavirus
NCI: National Cancer Institute
Pap: Papanicolaou
PSA: prostate-specific antigen
UIBE: University of International Business and Economics

Edited by R Kukafka; submitted 10.12.20; peer-reviewed by S Ma, S Lin, CY Lin, C Holmberg; comments to author 15.03.21; revised version received 10.05.21; accepted 14.06.21; published 27.08.21.

Please cite as:
Qin L, Zhang X, Wu A, Miser JS, Liu YL, Hsu JC, Shia BC, Ye L
Association Between Social Media Use and Cancer Screening Awareness and Behavior for People Without a Cancer Diagnosis: Matched Cohort Study
J Med Internet Res 2021;23(8):e26395
URL: https://www.jmir.org/2021/8/e26395
doi:10.2196/26395
PMID:34448708

©Lei Qin, Xiaomei Zhang, Anlin Wu, James S Miser, Yen-Lin Liu, Jason C Hsu, Ben-Chang Shia, Linglong Ye. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Distant Supervision for Mental Health Management in Social Media: Suicide Risk Classification System Development Study

Guanghui Fu[1*], MSc; Changwei Song[1*], BSc; Jianqiang Li[1], PhD; Yue Ma[2], PhD; Pan Chen[3], MSc; Ruiqian Wang[1], BSc; Bing Xiang Yang[3], PhD; Zhisheng Huang[4], PhD

[1]School of Software Engineering, Beijing University of Technology, Beijing, China
[2]Interdisciplinary Laboratory of Digital Sciences, Centre national de la recherche scientifique, Université Paris-Saclay, Orsay, France
[3]School of Health Sciences, Wuhan University, Wuhan, China
[4]Department of Artificial Intelligence, Vrije Universiteit Amsterdam, Amsterdam, Netherlands
[*]these authors contributed equally

Corresponding Author:
Bing Xiang Yang, PhD
School of Health Sciences
Wuhan University
Wuchang District
Wuhan, 430072
China
Phone: 86 15902731922
Email: 00009312@whu.wdu.cn

Abstract

Background: Web-based social media provides common people with a platform to express their emotions conveniently and anonymously. There have been nearly 2 million messages in a particular Chinese social media data source, and several thousands more are generated each day. Therefore, it has become impossible to analyze these messages manually. However, these messages have been identified as an important data source for the prevention of suicide related to depression disorder.

Objective: We proposed in this paper a distant supervision approach to developing a system that can automatically identify textual comments that are indicative of a high suicide risk.

Methods: To avoid expensive manual data annotations, we used a knowledge graph method to produce approximate annotations for distant supervision, which provided a basis for a deep learning architecture that was built and refined by interactions with psychology experts. There were three annotation levels, as follows: free annotations (zero cost), easy annotations (by psychology students), and hard annotations (by psychology experts).

Results: Our system was evaluated accordingly and showed that its performance at each level was promising. By combining our system with several important psychology features from user blogs, we obtained a precision of 80.75%, a recall of 75.41%, and an F1 score of 77.98% for the hardest test data.

Conclusions: In this paper, we proposed a distant supervision approach to develop an automatic system that can classify high and low suicide risk based on social media comments. The model can therefore provide volunteers with early warnings to prevent social media users from committing suicide.

(J Med Internet Res 2021;23(8):e26119) doi:10.2196/26119

KEYWORDS

deep learning; distant supervision; mental health; crisis prevention

Introduction

Background

Mental disorders have become a serious problem worldwide, with over 264 million people experiencing depression disorders [1]. A recent large-scale survey in China showed that the lifetime prevalence of depression and anxiety is 6.9% and 7.6%, respectively [2]. Depression is a leading cause of disability worldwide and contributes greatly to the global burden of diseases [3]. Suicide is the most serious consequence of depression [4]. The latest World Health Organization report showed that close to 800,000 people die by suicide every year. Furthermore, for each suicide, there are more than 20 suicide attempts. Suicides and suicide attempts have a ripple effect that impacts families, friends, colleagues, communities, and societies [5]. Stopping suicides before they are successful is a top priority. With the widespread popularity of the internet and the lack of immediate support in life, people are more inclined to express their emotions—even suicidal thoughts—in web-based communities, such as Weibo (Sina Corporation) and Twitter. Social media users favor these communities due to the anonymity and real-time advantages that they provide [6]. The data from web-based communities are huge in quantity, and it is difficult to analyze these data manually.

Traditional suicide risk assessment studies mainly conduct psychological tests, interviews, and questionnaires, which cost a lot of money. If computer technology can be used to assist suicide risk assessments, we can greatly improve the coverage and efficiency of screening and therefore reduce the number of suicide attempts. In recent years, many deep learning methods have been used for text sentiment analysis. However, these methods require large amounts of labeled data. With regard to the topic of our study, several thousands of comments are generated every day, but only a few (typically, less than 10) are indicative of high suicide risk. It is very time consuming to label so much data, and analyzing low and high suicide risk requires professional knowledge and special training.

In this context, we propose a distant supervision model to reduce the workload of domain experts. We integrated interesting scientific findings from the psychology field into our model. We developed a system that requires no manual annotations and takes into account the feedback of experts to better classify people's suicide risk levels based on the textual comments published on a Chinese social media platform. To avoid expensive manual data annotations, we used a knowledge graph method to produce approximate annotations, which provided a basis for building a deep learning model. The learning model was further refined by interactions with people with different experiences in psychology (beginners and experts) to generate our three data sets—the free annotation (zero cost), easy annotation (labeled by psychology students), and hard annotation (labeled by psychologists) data sets. We built the following three progressive models to fit these data sets: a bidirectional encoder representations from transformers (BERT)–based model, a fine-tuning model, and the psychology+ model. We obtained a precision of 80.75%, a recall of 75.41%, and an F1 score of 77.98% for the hard annotation data set, which was the hardest data set among the three data sets. This system reports on suicide risk assessments to volunteers from the Tree Hole Rescue Team [7] for web-based crisis prevention in China.

We first introduce the background of our research and discuss related work in the *Introduction* section. In the *Methods* section, we describe the task and introduce our workflow. In the *Results* section, we describe our data sets and a series of experimental results to prove the performance of our method. In the *Discussion* section, we summarize a series of open problems in this study and put forward some thoughts on future research directions. We summarize the whole paper in the final *Conclusion* section.

Related Work

Distant Supervision

Distant supervision is a method for using prior knowledge to generate noise labels (data containing wrong labels), which can help with avoiding a lot of manual labeling. In 2009, Mintz et al [8] first proposed the distant supervision method and used it for relation extraction. They generated a large amount of noise labels by mapping an existing knowledge base to rich unstructured text. Thus, they were able to train a good relationship extractor. Some sentiment classification methods also involve the distant supervision method. Go et al [9] proposed a method for automatically classifying sentiment in Twitter messages. They first used the emoticons in the text to generate labels and then used machine learning algorithms for classification. Purver and Battersby [10] and Suttles and Ide [11] used emoticons and hashtags or emojis in Twitter data to generate noise labels. Sahni et al [12] used half of a data set for comparisons with a baseline model and achieved a 2% to 3% higher accuracy than the baseline model. Camacho-Collados et al [13] proposed a simple, distant, supervision–based postprocessing step for using noisy, user-generated text to learn cross-lingual embeddings. They used English data to train classifiers for other languages. Purver and Battersby [10] used tag symbols and topic tags as weak tag training models. Without human intervention, the models were used to detect multiple types of emotions in Twitter messages. Mohammed et al [14] proposed a novel distant supervision technique for automatically gathering noisy topic category annotations from topically focused streams.

Most related research is based on the same annotation standard and eliminates the differences between different data through distance supervision. In our study, different levels of annotators (from nonprofessional to professional) provided different standards. Our method can be used with different levels of annotators (from basic algorithms to domain experts) to train models and obtain performance improvements to deal with real scenarios.

Deep Learning–Based Text Sentiment Analysis

Text sentiment analysis is the task of detecting sentiment information contained in text through a computer program. A basic task in sentiment analysis is classifying the polarity of a given text, that is, whether the expressed opinion in the text is positive, negative, or neutral. In advanced cases, polarity can refer to emotional states such as anger, sadness, and happiness. Sentiment analyses have been applied in marketing, customer service, and clinical medicine. Different from classical text sentiment analysis, our task of classifying high and low suicide risk on the basis of a given text was mainly based on whether users had decided a specific suicide method and a determinate suicide plan. Negative polarity and the emotional states of sadness and anger unnecessarily imply a high risk.

In recent years, many deep learning methods have been used for text sentiment analysis. Kim [15] used the word2vec method to convert sentences into sentence vectors and input the vectors into a convolutional neural network for sentiment analysis and question classification. Kalchbrenr et al [16] designed a dynamic convolutional neural network for binary and multi-class sentiment prediction. The model can handle input sentences of different lengths and can explicitly capture short and long relationships. Dieng et al [17] proposed a recurrent neural network for the sentiment analysis of movie reviews, which could capture the long-term dependency of text sentiment analysis. However, these methods all use word embedding for word vector representation, which cannot solve the problem of polysemy in reality. In 2018, Devlin et al [18] proposed a BERT-based, pretrained language model for dynamically obtaining encoded word vectors according to context. The BERT-based, pretrained language model uses a bidirectional transformer [19] as a feature extractor. Compared to traditional word embedding methods, such as Word2Vec [20-22] and Glove [23], the BERT method is better for capturing the representation of word and sentence levels. The BERT model is one of the most state-of-the-art pretraining models. Our final model was modified based on BERT, integrated the psychology features (Textbox 1) of users, and provided a more comprehensive suicide risk detection system.

Textbox 1. Psychology features and their definitions.

Self-description length

- The number of bytes that a user's homepage simply introduces to them

Weibo originality rate

- The ratio of the amount of original Weibo data to the total amount of Weibo data

Weibo link rate

- The ratio of the number of linked Weibo pages to the total number of Weibo pages

Weibo interaction rate

- The average number of mentions of other people in a user's Weibo account

Collective attention rate

- The average number of times each Weibo user uses first-person plural terms

Self-focus rate

- The average number of times each Weibo user uses first-person singular terms

Nighttime activity rate

- The ratio of the number of active users on Weibo from 10 PM to 6 AM to the total number of Weibo users

Social activity rate

- The ratio of the number of mutual attentions and the number of followers that a Weibo user has to those of other users

Willingness to express rate

- The ratio of the number of a user's Weibo comments to the number of their likes

Social support rate

- The ratio of the number of a user's Weibo comments to the total number of Weibo comments

Methods

Study Design

As previously mentioned, the information for classifying high and low suicide risk in our task is different from those used in the classical sentiment analysis tasks. In this study, to avoid the high cost of creating manually annotated data, we used a distant supervision approach that does not require manual annotations. Domain experts can use this approach to provide a small amount of annotations, which provides a basis for further improving a model by taking into account the experts' feedback. The flowchart of our method is shown in Figure 1.

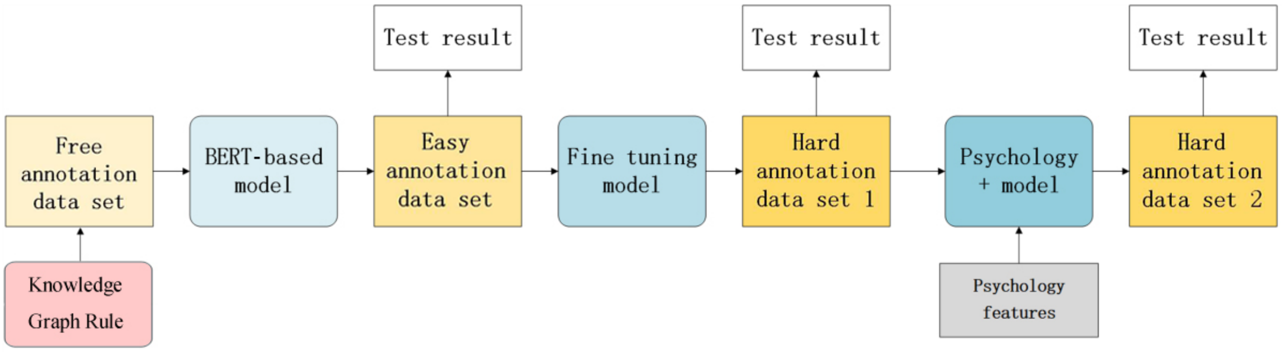
The method was divided into two parts—automated annotation generation (via the knowledge graph rules in Figure 1) and deep learning–based classification (the three models in Figure 1).

For automated annotation generation, a set of semantic rules was constructed based on an ontology for the field of crisis prevention to generate the free annotation data set. These automated (possibly erroneous) annotations were then used to supervise the deep learning models.

We then build a BERT-based model based on the free annotation data set. This model was used to classify new texts that were further corrected by psychology students to generate the easy annotation data set. This data set had comparable amounts of high and low crisis risk data. It should be noted that without the assistance of a computer algorithm, it would have been a massive challenge for humans to provide such a balanced data set because the percentage of high-risk messages was quite low, as mentioned above. We used the easy annotation data set to fine-tune this basic learning model and develop the fine-tuning model, which took into account the knowledge in the easy annotations. In parallel, a psychology expert assisted with providing the hard annotation data set, which is of much smaller size than the easy annotation data set due to its cost. Finally, we improved our model by using the hard annotation data set and extra psychology features to obtain the final psychology+ model.

Our models continually fitted 3 data sets. These three data sets were labeled by psychology practitioners of different levels. As the standards for the labeling process gradually became stricter, the models became more accurate. The final model (psychology+ model) combined the prior knowledge obtained from the precedent models and fused prior domain knowledge. Improved performance was achieved under the premise of using a small amount of manually annotated data.

Figure 1. Flowchart of our method. As the labeling process became stricter, we continued to improve our model's performance. BERT: bidirectional encoder representations from transformers.



Task Definition

Given a textual sentence (*s*) published on a web-based media platform by a user (*u*), our task was to predict if the user (*u*) has a high suicide risk. If this is the case, a crisis intervention will be provided to save the user's life. This is a challenging task, and even a person with professional knowledge has to be careful to avoid mistakes. Furthermore, the huge amount of comments produced daily makes manual analysis impossible and costly.

With regard to the definitions of high and low risk, Huang et al [24] proposed a grading standard based on the certainty of planned suicide methods and the urgency of the action, thereby obtaining 10 risk levels, as shown in Textbox 2. We defined level 6 and above as "high risk" and the rest as "low risk." Furthermore, social media may contain a lot of irrelevant comments. Therefore, we created level 0 and defined it as "no risk."

Textbox 2. Suicide risk levels and judgement criteria.

Level 10

- Suicide may be ongoing

Level 9

- Suicide method has been determined
- Person may commit suicide in the near future

Level 8

- Suicide has been planned
- Date of suicide has been roughly determined

Level 7

- Suicide method has been determined but not the suicide date

Level 6

- Suicide method is planned
- Suicide method is unknown

Level 5

- Strong suicide desire
- Unclear mode of suicide

Level 4

- Suicide wish has been expressed
- Suicide method and plan are unclear

Level 3

- Strong survival pain
- No expression of suicidal wish

Level 2

- The pain of survival has been clearly expressed
- No suicidal desire

Level 1

- Expression of survival pain
- Expression of suicide desire

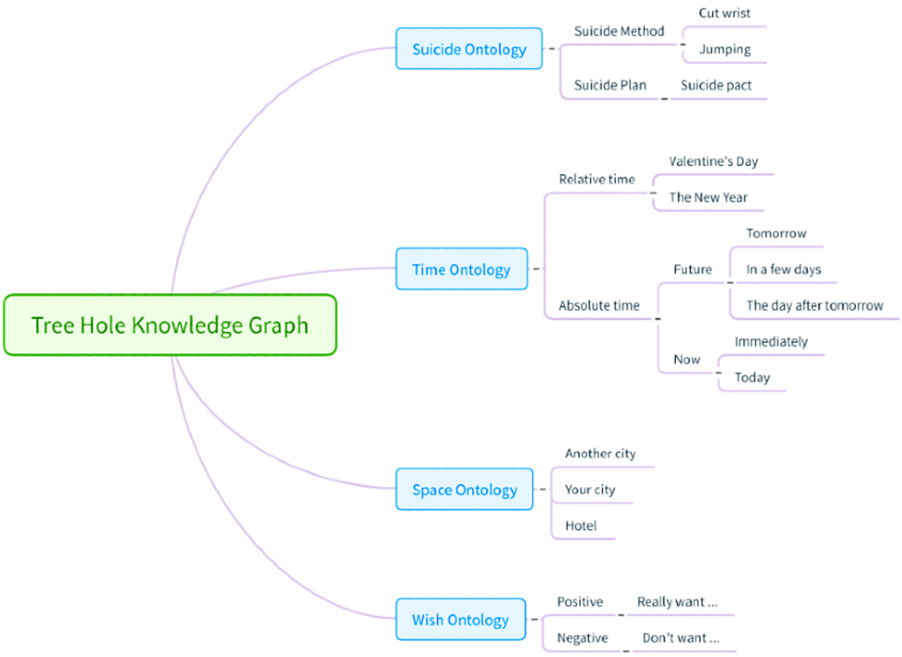
Level 0

- No expression of survival pain

Knowledge Graph and Reasoning Rules

We created a knowledge graph that was used to construct a free annotation data set. The Tree Hole knowledge graph contains four independent ontologies—the suicide ontology, time ontology, space ontology, and wish ontology. The suicide ontology consists of two major categories—suicide methods (eg, cutting the wrist and burning charcoal) and suicide plans (eg, buying drugs and meeting with suicide partners). The time ontology covers absolute time concepts, such as calendar days and holidays, and relative time concepts, such as the present, future, and past. The space ontology describes related concepts of spatial geography. The wish ontology was mainly used to analyze the subjective suicidal wishes of a specific group of people and to exclude them from people without subjective suicidal wishes. An excerpt of the Tree Hole knowledge graph can be found in Figure 2.

Figure 2. An excerpt of the Tree Hole knowledge graph.


Tree Hole Knowledge Graph
Suicide Ontology
Suicide Method
Cut wrist
Jumping
Suicide Plan
Suicide pact
Time Ontology
Relative time
Valentine's Day
The New Year
Absolute time
Future
Tomorrow
In a few days
The day after tomorrow
Now
Immediately
Today
Space Ontology
Another city
Your city
Hotel
Wish Ontology
Positive
Really want ...
Negative
Don't want ...

Once the knowledge graph was created, we constructed prolog rules based on the definite clause grammar (DCG) and DCG transformation rules. To take into account the domain knowledge for reasoning and determining the risk level, the DCG rules integrate the relevant conceptual information from the Tree Hole knowledge graph. For example, the definition for suicide risk level 8 is that the suicide plan has been determined and the suicide date has been roughly determined. This can be formally described by the logic program rules in Textbox 3, which use two predicates—*rdfsSubClassOf* and *uninterestedText*. The predicate *rdfsSubClassOf* was used to decide if a textual fragment mentions some sort of suicide method (or time) concept of *suicideOntology* (or *timeOntology*) in the Tree Hole knowledge graph, and *uninterestedText* was used to verify that such a mention was not negated by a negative expression in the text. More details can be found in other studies [24,25].

Textbox 3. Logic program rules for suicide risk level 8.

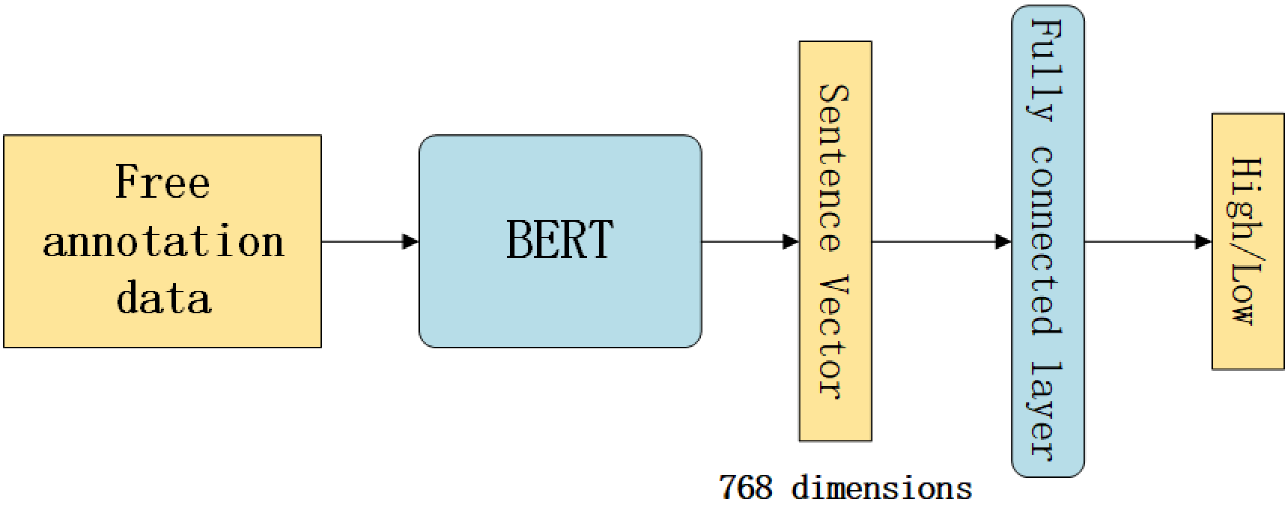
Program rules

- statement(suicideRisk(8,[Plan, Time]))→ uninterestedText(_L1), rdfsSubclassOf(Time,future,timeOntology), uninterestedText(_L2), rdfsSubclassOf(Plan, suicidePlan, suicideOntology), uninterestedText(_L3)
- statement(suicideRisk(8, [Plan, Time]))→ uninterestedText(_L1), rdfsSubclassOf(Plan, suicidePlan, suicideOntology), uninterestedText(_L2), rdfsSubclassOf(Time, future, timeOntology), uninterestedText(_L3).

BERT-Based Model

We used the data set generated by Tree Hole knowledge graph method to build the BERT-based model. We used BERT to obtain sentence vectors from free annotation data. The size of each sentence vector was 768 dimensions. We added a dropout function to this sentence vector to avoid overfitting. Afterward, we added a fully connected layer to classify comments that were indicative of high and low suicide probabilities. We used the sigmoid function as the activation function of the output layer. The parameters of the BERT layer and the fully connected layer participated in the training at the same time. The architecture of model 1 is shown in Figure 3.

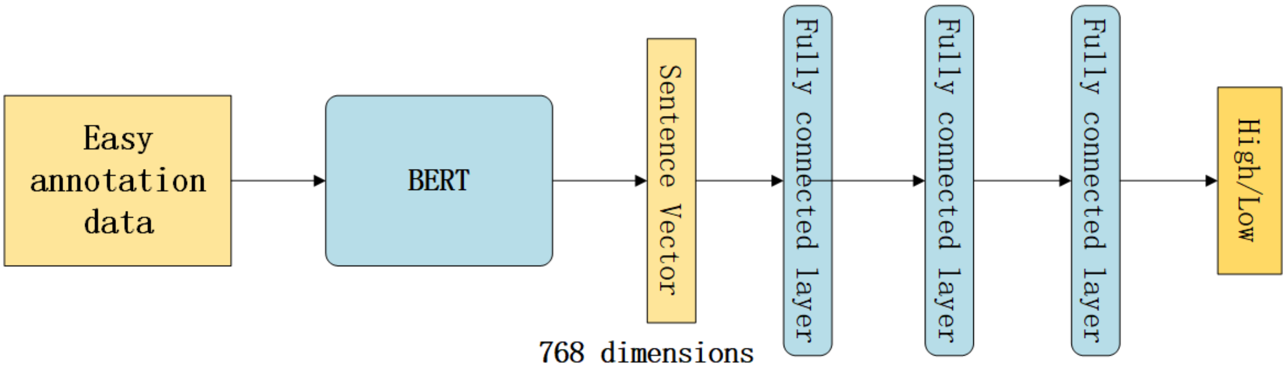
Figure 3. Architecture of our BERT-based model. BERT: bidirectional encoder representations from transformers .



Fine-Tuning Model

We used BERT-trained (the first model) to obtain 768D sentence vectors from the easy annotation data set. We added 3 fully connected layers to the sentence vectors. The input and output of the first 2 fully connected layers had 768 dimensions, and we used the ReLU function [26] as their activation function. We also added a dropout function and batch normalization function [27] after each activation function to prevent overfitting. We used the sigmoid function as the activation function of the output layer. Different from the first model, the fine-tuning model did not train the BERT layer but only trained the previous fully connected layer. The structure of model 2 is shown in Figure 4.

Figure 4. Architecture of our fine-tuning model. BERT: bidirectional encoder representations from transformers.

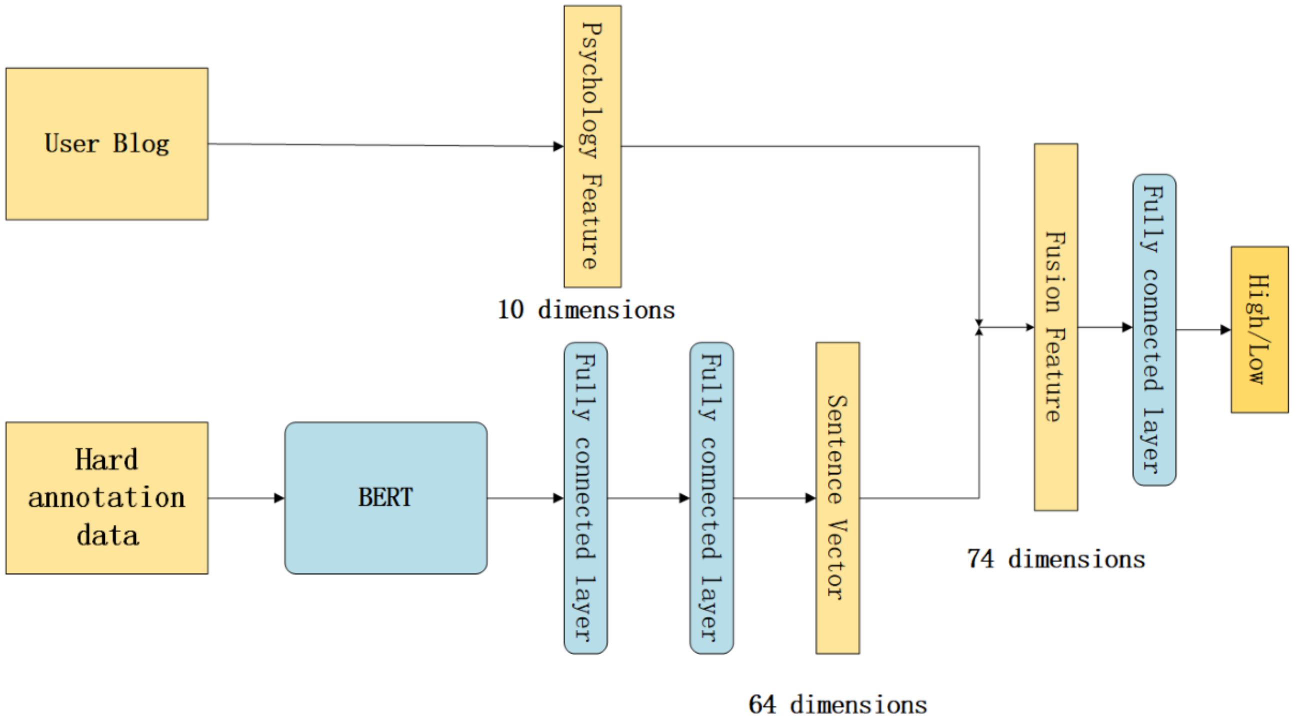


Psychology+ Model

This model combined deep learning and psychology features. According to a psychology study, people with a high suicide risk have a higher degree of self-concern than those with a low suicide risk, that is, they may be more focused on themselves rather than their surroundings [28]. Social network behaviors can reflect individuals' personality traits [29]. Gosling et al [30] found that people with extraverted personality traits use social networks more frequently, have higher engagement, and are more active compared to introverts. To take advantage of these scientific findings, we built the psychology+ model, as described below.

We analyzed the 10 psychology features defined in Textbox 1. According to psychological research, people with a high and low suicide risk have significantly different behaviors with regard to these 10D features [28-32]. We further analyzed and obtained Weibo users' psychology features and integrated them into our final model. These 10 features were represented as a 10D vector. Afterward, we used the BERT-trained model (the first model) to obtain 768D sentence vectors from hard annotation data. We used 2 fully connected layers to reduce the sentence vector to 64 dimensions. Our choice of using a 64D vector was based on experiments. We experimented with 256, 128, 64, 32 dimensions, and the best results were obtained with 64 dimensions. We combined the 64D sentence vector and 10D psychological feature vector to form a 74D vector. Finally, we used the fully connected layer as a classifier and the sigmoid function as the activation function to classify high and low suicide risk. This model structure is shown in Figure 5.

Figure 5. Architecture of our psychology+ model. BERT: bidirectional encoder representations from transformers.



Results

Data Set

In this study, data were obtained from the comments of the Zoufan Weibo account [33]. In addition, we performed the following preprocessing methods for the data sets: the removal of duplicate data, the removal of meaningless emojis, and the removal of too short sentences (≤5 words). As previously explained, we constructed the following four data sets (our annotation data sets can be made available upon request for research purposes):

- the free annotation data set, which was generated by using knowledge graph method
- the easy annotation data set, which was labeled by a psychology student
- hard annotation data sets 1 and 2, which were labeled by a psychologist

These data sets were used to train and test our models. The data distribution can be seen in Table 1.

For the psychology+ model, the psychology features were extracted from each user's Weibo page [34].

Table 1. Comment distributions in each data set.

Data set	Comments indicating a high suicide risk, n	Comments indicating a low suicide risk, n
Free annotation data set	3630	3220
Easy annotation data set	6659	8657
Hard annotation data set 1	813	645
Hard annotation data set 2	599	601

Evaluation

We evaluated the three learning models that were built based on the automatically generated annotation data set. We found that the free and easy annotation data sets resulted in a simple classification task that could be solved well by the BERT-based model and the fine-tuning model, respectively. However, the hard annotation data set resulted in a much harder task for which our psychology+ model could achieve a promising performance.

Experimental Setup

For the basic BERT-based model, we used the pretrained Chinese model released by Google. The model uses a 12-layer transformer with about 110 million parameters. The optimizer uses the Adam method [35], and the learning rate was set to 0.00002. The maximum length of inputted characters is 128. The batch sizes of the training set, validation set, and test set were 8, 16, and 16, respectively. At the fully connected layer, the dropout was set to 0.1. We built our model by using a GTX 2080Ti graphics processing unit, and the deep learning framework we used was PyTorch [36]. In the evaluation experiments, we used F1 scores, precision, recall, and the accuracy of high-risk classifications as the evaluation metrics. The F1 score is the harmonic mean of precision and recall, and it was the most important evaluation metric in our study.

Evaluation of the BERT-Based Model

We performed fivefold cross-validation for 6850 comments (3630 comments indicating a high suicide risk and 3220 indicating a low suicide risk; Table 1) and thus obtained 5480 comments for training and 1370 comments for testing. We report in Table 2 the evaluation results of the model that was based on the free annotation data set and the easy annotation data set.

We found that the values of all evaluations metrics for the free annotation data set were higher than 0.98. This shows that this simple model can simulate the knowledge graph–based approach well. The model's performance based on the easy annotation data set was lower, particularly in terms of precision (0.9899 vs 0.8367), indicating that the model needs further improvement to analyze behavior just as well as psychology students for the annotation task.

Table 2. Bidirectional encoder representations from transformers–based model.

Data set	F1 score	Precision	Recall	Accuracy
Free annotation data set	0.9864	0.9899	0.9829	0.9862
Easy annotation data set	0.9111	0.8367	0.9998	0.9151

Evaluation of the Fine-Tuning Model

According to the fivefold cross-validation, we separated 15,316 comments (6659 comments indicating a high suicide risk and 8657 indicating a low suicide risk; Table 1) into the training set or the test set, which contained 12,252 and 3064 comments, respectively. We also evaluated the performance of the model based on hard annotation data sets 1 and 2.

The results can be seen in Table 3. Clearly, by fine-tuning the model, the performance of the model improved (compared to the BERT-based model) in terms of F1 score (0.9111 vs 0.9214), precision (0.8367 vs 0.9241), and accuracy (0.9151 vs 0.9218). However, the recall value dropped a little bit (0.9998 vs 0.9282).

In contrast to the performance of the model based on the easy annotation data set, the results for the 2 hard annotation data sets were unsatisfactory (<0.6 in most cases). This meant that the model needed further improvement. Intuitively, we could have performed a similar fine-tuning process with the hard annotations. However, as seen in Table 1, the hard annotation data sets were much smaller than the easy annotation data set. This was due to the high cost of the hard annotations. Therefore, we created the psychology+ model by taking into account certain psychological knowledge in order to avoid the requirement of large amounts of annotated data.

Table 3. Fine-tuning model.

Data set	F1 score	Precision	Recall	Accuracy
Easy annotation data set	0.9214	0.9241	0.9282	0.9218
Hard annotation data set 1	0.7281	0.5815	0.9734	0.5942
Hard annotation data set 2	0.6753	0.5131	0.9877	0.5239

Evaluation of the Psychology+ Model

According to fivefold cross-validation, we separated 1458 comments (813 comments indicating a high suicide risk and 645 indicating a low suicide risk from hard annotation data set 1; Table 1) into the training set or the validation set, which contained 1167 and 291 comments, respectively. Hard annotation data set 2 was reserved mainly for testing. We report the evaluation results for these two hard annotation data sets in Table 4. We obtained an accuracy of 78.68% for hard annotation data set 2. Compared to the fine-tuning model, the F1 score of the psychology+ model based on hard annotation data sets 1 and 2 improved by 7.88% and 10.45%, respectively.

Table 4. Psychology+ model.

Data set	F1 score	Precision	Recall	Accuracy
Hard annotation data set 1	0.8069	0.8067	0.8072	0.8105
Hard annotation data set 2	0.7798	0.8075	0.7541	0.7868

Discussion

Principal Findings

Summary of Findings

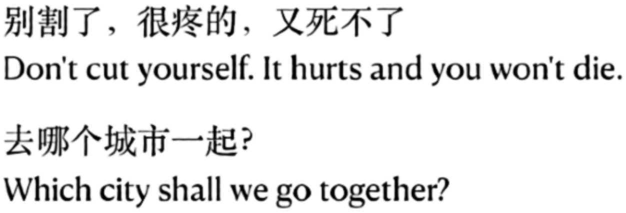
In this study, we examined three types of annotation data sets—the free annotation, easy annotation, and hard annotation data sets. As seen in Table 2 and Table 3, free annotations and easy annotations can be generated by deep learning models of high quality. Even though we achieved promising results from the hard annotation data sets, the quality of the results is still imperfect. Indeed, our task was exposed to some challenges due to several characteristics of the data.

We consider the F1 score to be the most important evaluation metric. In follow-up work, we will invite more volunteers to manually evaluate our model based on actual situations.

Obscure Expression

With regard to the two sentences in Figure 6, without considering the context of the task, neither of these two statements are indicative of a high risk of suicide. However, due to the nature of our data source, both sentences were classified as high risk by psychological experts. The first sentence expresses the result of someone cutting their wrist, but in terms of our context, the sentence was used to comment on a suicide attempt (cutting wrist), indicating that the attempt has been taken. Therefore, a prevention intervention should be provided. The second sentence is about making an appointment with other people to die together. In the context of our data source, it can be considered that the date of the suicide was about to be determined, resulting in a high risk level of between 7 and 8.

Figure 6. Examples of comments that express a high suicide risk in an obscure way.



别割了，很疼的，又死不了

Don't cut yourself. It hurts and you won't die.

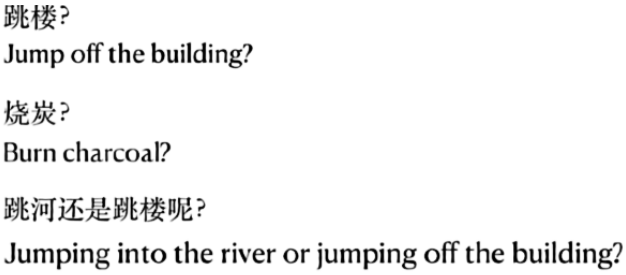
去哪个城市一起?

Which city shall we go together?

Ultrashort Text

Short texts can also indicate a high suicide risk. For example, for the three short sentences in Figure 7, the first two are indicative of a low suicide risk, but the third is indicative of a high suicide risk. The is because the first two expressions indicate that the plan is still under consideration. However, the third indicates that the plan has been restricted to two possible suicide methods. Such ultrashort texts are difficult to classify, and successful classification needs the combination of more specific context and domain knowledge.

Figure 7. Ultrashort texts that are indicative of different suicide risk levels.



跳楼?

Jump off the building?

烧炭?

Burn charcoal?

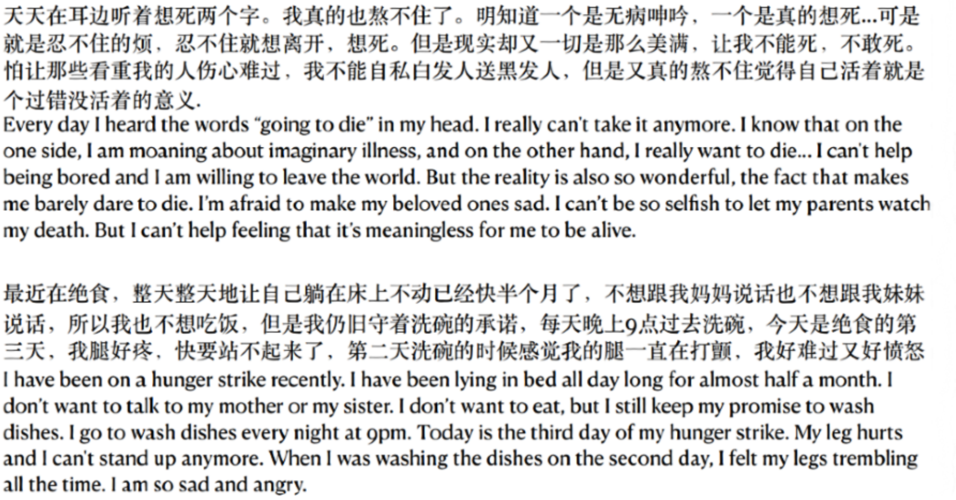
跳河还是跳楼呢?

Jumping into the river or jumping off the building?

Long Text With Contradictions

Some people express their feelings by telling their own stories, as illustrated by the examples given in Figure 8. These long, contradictory, redundant, and possibly incomprehensible expressions and sentences also make classification difficult.

Figure 8. Examples of long texts for which the suicide intention level is hard to capture.



天天在耳边听着想死两个字。我真的也熬不住了。明知道一个是无病呻吟，一个是真的想死...可是就是忍不住的烦，忍不住就想离开，想死。但是现实却又一切是那么美满，让我不能死，不敢死。怕让那些看重我的人伤心难过，我不能自私白发人送黑发人，但是又真的熬不住觉得自己活着就是个过错没活着的意义.

Every day I heard the words "going to die" in my head. I really can't take it anymore. I know that on the one side, I am moaning about imaginary illness, and on the other hand, I really want to die... I can't help being bored and I am willing to leave the world. But the reality is also so wonderful, the fact that makes me barely dare to die. I'm afraid to make my beloved ones sad. I can't be so selfish to let my parents watch my death. But I can't help feeling that it's meaningless for me to be alive.

最近在绝食，整天整天地让自己躺在床上不动已经快半个月了，不想跟我妈妈说话也不想跟我妹妹说话，所以我也不想吃饭，但是我仍旧守着洗碗的承诺，每天晚上9点过去洗碗，今天是绝食的第三天，我腿好疼，快要站不起来了，第二天洗碗的时候感觉我的腿一直在打颤，我好难过又好愤怒

I have been on a hunger strike recently. I have been lying in bed all day long for almost half a month. I don't want to talk to my mother or my sister. I don't want to eat, but I still keep my promise to wash dishes. I go to wash dishes every night at 9pm. Today is the third day of my hunger strike. My leg hurts and I can't stand up anymore. When I was washing the dishes on the second day, I felt my legs trembling all the time. I am so sad and angry.

Conclusion

In this paper, we proposed a distant supervision approach to develop an automatic system that can classify high and low suicide risk based on social media comments. We constructed 3 data sets of different levels (free, easy, and hard) via interactions with psychologists who were assisted by our models.

Although a deep learning model has excellent performance in different domains, it requires a lot of annotations to train a reliable model. Per our study, ordinary people cannot accurately label data. Only people who have been professional trained can accurately classify the suicide risk expressed in comments in accordance with standard methods. This makes it difficult to obtain large-scale annotations. Per our processing steps, we first used a basic algorithm to generate annotations for training the baseline model. Afterward, we invited people with different psychology knowledge levels to provide a small number of annotations. Then, based on domain knowledge, we extracted users' multidimensional psychological features and integrated them into our final model (the psychology+ model). Our operating steps greatly improved the efficiency of our model. Only 1458 professional labels were required to train a model that could analyze real situations. It would have been impossible to train a reliable model with just over 1000 data points if we did not conduct the previous steps.

In future work, we will combine actual work experiences and cooperate with psychologists to propose more suitable suicide classification standards and provide immediate warnings for upcoming emergencies. Although our model could meet actual standard requirements, its run time was relatively slow (274 comments/minute) due to the large number of model parameters. Further, although the model's efficiency could meet people's daily needs, we still hope to develop a more lightweight model for dealing with certain data produced in special situations, such as short-term, large-scale comments.

Acknowledgments

This study is supported by the National Natural Science Foundation of China (No. 72174152) and the Project of Humanities and Social Sciences of the Ministry of Education in China (The Proactive Levelled Intervention for Social Network Users' Emotional Crisis—an Automatic Crisis Balance Analysis Model, 20YJCZH204).

Conflicts of Interest

None declared.

References

1. Depression. World Health Organization. 2020 Jan 30. URL: https://www.who.int/news-room/fact-sheets/detail/depression [accessed 2020-01-30]
2. Huang Y, Wang Y, Wang H, Liu Z, Yu X, Yan J, et al. Prevalence of mental disorders in China: a cross-sectional epidemiological study. Lancet Psychiatry 2019 Mar;6(3):211-224. [doi: 10.1016/S2215-0366(18)30511-X] [Medline: 30792114]
3. World Health Organization. Mental Health Action Plan 2013 - 2020. Geneva, Switzerland: World Health Organization; 2013.
4. Li H, Luo X, Ke X, Dai Q, Zheng W, Zhang C, et al. Major depressive disorder and suicide risk among adult outpatients at several general hospitals in a Chinese Han population. PLoS One 2017 Oct 10;12(10):e0186143. [doi: 10.1371/journal.pone.0186143] [Medline: 29016669]
5. Suicide. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/suicide [accessed 2019-09-02]
6. Chen P, Qian Y, Huang Z, Zhao C, Liu Z, Yang B, et al. Negative emotional characteristics of weibo ”tree hole” users. Zhongguo Xinliweisheng Zazhi 2020;5:437-444.
7. Tree Hole Rescue Team. Weibo. URL: https://weibo.com/u/5133305819 [accessed 2021-07-06]
8. Mintz M, Bills S, Snow R, Jurafsky D. Distant supervision for relation extraction without labeled data. 2009 Presented at: Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP; August 2-7, 2009; Suntec, Singapore p. 1003-1011. [doi: 10.3115/1690219.1690287]
9. Go A, Bhayani R, Huang L. Twitter sentiment classification using distant supervision. Stanford University. URL: https://www-cs.stanford.edu/people/alecmgo/papers/TwitterDistantSupervision09.pdf [accessed 2021-06-29]
10. Purver M, Battersby S. Experimenting with distant supervision for emotion classification. 2012 Presented at: 13th Conference of the European Chapter of the Association for Computational Linguistics; April 23-27, 2012; Avignon, France p. 482-491.
11. Suttles J, Ide N. Distant supervision for emotion classification with discrete binary values. 2013 Presented at: 14th International Conference on Intelligent Text Processing and Computational Linguistics; March 24-30, 2013; Samos, Greece p. 121-136. [doi: 10.1007/978-3-642-37256-8_11]
12. Sahni T, Chandak C, Chedeti NR, Singh M. Efficient Twitter sentiment classification using subjective distant supervision. 2017 Presented at: 2017 9th International Conference on Communication Systems and Networks (COMSNETS); January 4-8, 2017; Bengaluru, India. [doi: 10.1109/comsnets.2017.7945451]
13. Camacho-Collados J, Doval Y, Martínez-Cámara E, Espinosa-Anke L, Barbieri F, Schockaert S. Learning cross-lingual word embeddings from Twitter via distant supervision. arXiv. Preprint posted online on March 31, 2020 [FREE Full text]
14. Mohammed S, Ghelani N, Lin J. Distant supervision for topic classification of Tweets in curated streams. arXiv. Preprint posted online on April 22, 2017 [FREE Full text]

15. Kim Y. Convolutional neural networks for sentence classification. arXiv. Preprint posted online on September 3, 2014 [FREE Full text]
16. Kalchbrenner N, Grefenstette E, Blunsom P. A convolutional neural network for modelling sentences. arXiv. Preprint posted online on April 8, 2014 [FREE Full text]
17. Dieng AB, Wang C, Gao J, Paisley J. TopicRNN: A recurrent neural network with long-range semantic dependency. arXiv. Preprint posted online on November 5, 2016 [FREE Full text]
18. Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv. Preprint posted online on October 11, 2018 [FREE Full text]
19. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: Advances in neural information processing systems. 2017 Presented at: 31st International Conference on Neural Information Processing Systems; December 4, 2017; Long Beach, California, USA p. 6000-6010.
20. Bengio Y, Ducharme R, Vincent P, Jauvin C. A neural probabilistic language model. J Mach Learn Res 2003 Feb;3:1137-1155 [FREE Full text] [doi: 10.5555/944919.944966]
21. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. arXiv. Preprint posted online on September 7, 2013 [FREE Full text]
22. Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed representations of words and phrases and their compositionality. arXiv. Preprint posted online on October 16, 2013 [FREE Full text]
23. Pennington J, Socher R, Manning C. Glove: Global vectors for word representation. 2014 Presented at: 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP); October 2014; Doha, Qatar p. 1532-1543. [doi: 10.3115/v1/d14-1162]
24. Huang Z, Hu Q, Gu J, Yang J, Feng Y, Wang G. Web-based intelligent agents for suicide monitoring and early warning. China Digital Medicine 2019;14(3):2-6.
25. Yang F, Huang Z, Yang B, Ruan J, Nie W, Fang S. Analysis of suicidal ideation patterns weibo "tree hole" users based on artificial intelligence technology. Journal of Nursing Science 2019;34(24):42-45 [FREE Full text]
26. Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks. 2011 Presented at: 14th International Conference on Artificial Intelligence and Statistics (AISTATS); April 11-13, 2011; Ft. Lauderdale, FL, USA URL: http://proceedings.mlr.press/v15/glorot11a/glorot11a.pdf
27. Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv. Preprint posted online on February 11, 2015 [FREE Full text]
28. Guan L, Hao B, Cheng QJ, Ye Z, Zhu T. Behavioral and linguistic characteristics of microblog users with various suicide ideation level: An explanatory study. Chinese Journal of Public Health 2015;31(3):349-352.
29. Ryan T, Xenos S. Who uses Facebook? An investigation into the relationship between the Big Five, shyness, narcissism, loneliness, and Facebook usage. Comput Human Behav 2011 Sep;27(5):1658-1664. [doi: 10.1016/j.chb.2011.02.004]
30. Gosling SD, Augustine AA, Vazire S, Holtzman N, Gaddis S. Manifestations of personality in online social networks: self-reported Facebook-related behaviors and observable profile information. Cyberpsychol Behav Soc Netw 2011 Sep;14(9):483-488 [FREE Full text] [doi: 10.1089/cyber.2010.0087] [Medline: 21254929]
31. Cash SJ, Thelwall M, Peck SN, Ferrell JZ, Bridge JA. Adolescent suicide statements on MySpace. Cyberpsychol Behav Soc Netw 2013 Mar;16(3):166-174. [doi: 10.1089/cyber.2012.0098] [Medline: 23374167]
32. Spates K, Ye X, Johnson A. "I just might kill myself": Suicide expressions on Twitter. Death Stud 2020;44(3):189-194. [doi: 10.1080/07481187.2018.1531085] [Medline: 30556789]
33. Zoufan. Weibo. URL: https://www.weibo.com/xiaofan116?is_all=1 [accessed 2021-07-06]
34. Weibo. Weibo. URL: https://www.weibo.com/ [accessed 2021-07-06]
35. Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv. Preprint posted online on December 22, 2014 [FREE Full text]
36. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. PyTorch: An imperative style, high-performance deep learning library. 2019 Presented at: Advances in Neural Information Processing Systems 32 (NeurIPS 2019); December 8-14, 2019; Vancouver, BC, Canada.

Abbreviations

BERT: bidirectional encoder representations from transformers
DCG: definite clause grammar

Edited by R Kukafka; submitted 28.11.20; peer-reviewed by X Xiu, A Teles; comments to author 26.01.21; revised version received 24.03.21; accepted 04.05.21; published 26.08.21.

Please cite as:
Fu G, Song C, Li J, Ma Y, Chen P, Wang R, Yang BX, Huang Z
Distant Supervision for Mental Health Management in Social Media: Suicide Risk Classification System Development Study
J Med Internet Res 2021;23(8):e26119
URL: https://www.jmir.org/2021/8/e26119
doi:10.2196/26119
PMID:34435964

©Guanghui Fu, Changwei Song, Jianqiang Li, Yue Ma, Pan Chen, Ruiqian Wang, Bing Xiang Yang, Zhisheng Huang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Factors That Help and Hinder the Implementation of Digital Depression Prevention Programs: School-Based Cross-sectional Study

Joanne R Beames[1], MPsych, PhD; Lara Johnston[1], MPH; Bridianne O'Dea[1], PhD; Michelle Torok[1], PhD; Helen Christensen[1], MPsychol, PhD; Katherine M Boydell[1], PhD; Aliza Werner-Seidler[1], MPsych, PhD

Black Dog Institute, University of New South Wales, Randwick, Australia

Corresponding Author:
Joanne R Beames, MPsych, PhD
Black Dog Institute
University of New South Wales
Hospital Road
Randwick, 2031
Australia
Phone: 61 2 9382 ext 8776
Email: j.beames@blackdog.org.au

Abstract

Background: Digital prevention programs that are delivered in a school environment can inoculate young people against depression. However, little is known about the school-based factors that help and hinder the implementation of these programs. Staff members are integral for supporting mental health programs in schools and are likely to have a wealth of expertise and knowledge about the factors that affect implementation.

Objective: The primary objective of this study was to explore the barriers and facilitators to implementing a digital depression prevention program in Australian secondary schools with teachers, counselors, and principals. The secondary objective was to explore variations in these factors across different school contexts, including the school type (government or nongovernment), location (capital city, regional/or rural areas), and socioeconomic status (SES) (low, medium, high).

Methods: This quantitative cross-sectional survey study assessed the barriers and facilitators to implementing a hypothetical digital prevention program in Australian schools. The survey was taken by 97 teachers (average age 38.3 years), 93 counselors (average age 39.5 years), and 11 principals (average age 50.9 years) across Australia between November 2017 and July 2018.

Results: A range of barriers and facilitators relating to logistics and resources, staff support, and program factors were endorsed by the surveyed staff. Consistent with prior research, common barriers included a lack of time and resources (ie, staff and rooms). These barriers were particularly evident in government, rural/regional, and low socioeconomic schools. Other barriers were specific to digital delivery, including privacy issues and a lack of clarity around staff roles and responsibilities. Facilitators included upskilling staff through training, embedding the program into the curriculum, and other program factors including universal delivery, screening of students' mental health, and clear referral pathways. Knowledge about the program efficacy was also perceived as important by a large proportion of the respondents.

Conclusions: The digital depression prevention program was perceived as suitable for use within different schools in Australia, although certain factors need to be considered to enable effective implementation. Logistics and resources, support, and program factors were identified as particularly important for school-based implementation. To maximize the effectiveness in delivering digital programs, implementation may need to be tailored to the staff roles and school types.

(J Med Internet Res 2021;23(8):e26223) doi:10.2196/26223

KEYWORDS

secondary school; depression; prevention; digital; barrier; facilitator; teacher; counselor; principal; student

Introduction

Depression is a debilitating problem in young people. An earlier onset is associated with a more severe clinical course and greater risk of recurrence, impairments in academic performance, social problems, and increased risk of comorbid physical and mental health problems [1,2]. Access to evidence-based treatment for depression is often limited, and even among those who do obtain access, relapse rates remain high [3]. Therefore, depression is a prime target for prevention. To increase access to and effectiveness of prevention efforts, a proactive approach is needed whereby programs are delivered in contexts frequented by young people using methods in which they are already engaged. One way to increase access and effectiveness is by delivering psychological prevention programs in schools.

Schools are ideally placed to address the mental health problems of young people. Young people spend much of their waking time at school and have reported a greater willingness to access mental health services at school than in other settings [4,5]. Among young people who do receive mental health care services, more than 50% have indicated that the first access was driven by their school [4]. Schools in many countries, including Australia, Canada, and the United Kingdom, typically have designated mental health and well-being staff (eg, counselors, psychologists, or well-being officers). Although there are variations across schools, these staff members are well positioned to support mental health programs and create meaningful impact for students.

School-based depression prevention programs are already available. Two meta-analyses have found that school-based prevention programs had a small preventive effect on depressive symptoms, regardless of whether the program was delivered to all students (ie, universal) or to a targeted sample [6,7]. Most of the studies involved programs that were delivered in person, an approach that places high demand on the already limited resources in schools (eg, staff availability, infrastructure, time, financial cost). Only two studies from this review reported on programs that were delivered online via a website [8,9]. More recent randomized controlled trials have investigated the effectiveness of online or telephone-based prevention programs delivered in schools, with mixed results [10-13]. Given that there is some indication of the effectiveness of school-based digital programs, using technology to deliver prevention programs may be a promising way forward in addressing student depression in schools.

Barriers and Facilitators to Delivering Mental Health Programs in Schools

Digital Programs

Little is known about the factors affecting the implementation of digital mental health programs in schools. Research has typically focused on the factors that generally affect the acceptability of and engagement with digital mental health programs, or on settings other than schools [14,15]. Schools are complex environments for delivering digital mental health programs, and many factors will affect their implementation and uptake by students. Given that school staff and school leaders decide what programs are implemented, and are often involved in supporting program delivery, it is important to consider their perspectives about what will work in their schools. Research investigating the barriers and facilitators of specific digital mental health programs as perceived by school staff is currently lacking [16], with more focus on face-to-face delivery methods.

Face-to-Face Programs

Studies have identified school-specific factors that affect implementation of face-to-face evidence-based mental health programs. One early qualitative study involved interviews with the developers of multiple evidence-based school mental health interventions [17]. Results identified the following seven factors as being important for effective implementation: (1) principal and other administrative support; (2) teacher support; (3) adequate financial resources; (4) high-quality training and consultation to ensure fidelity to the intervention components; (5) alignment of the intervention with the philosophy, culture, and approach to mental health of the schools; (6) visibility of the intervention outcomes to key stakeholders; and (7) adequate ways to address turnover of the staff involved in program support and delivery. Another study was conducted based on these results by incorporating the perspectives of program directors and clinicians working within schools [18]. In this study, Langley and colleagues found that the four main barriers to implementation included competing responsibilities, parent engagement, logistics, and support from administrators and teachers. Facilitators included having a professional network and being able to consult others about the program, adequate funding, and perceptions that the program would be easy to implement. Other survey studies involving school staff (eg, headteachers, teachers, counselors) in middle and secondary schools have obtained converging results [19-26]. School Mental Health ASSIST, a provincial implementation support team in Ontario that aims to build school capacity to use evidence-based strategies and services, has also identified a “top 10 list” of factors that influence implementation of student mental well-being practices. These factors include commitment, leadership team, assessment of the (initial) capacity and resources, and professional learning [27]. This literature underscores the need to consider factors unique to the school environment that influence the delivery of mental health programs, rather than relying on the more generic implementation literature.

Current Study

To the best of our knowledge, no studies have evaluated the barriers and facilitators to implementing a school-based digitally delivered depression prevention program for students as perceived by school staff. Although insights can be drawn from studies evaluating the implementation of evidence-based face-to-face interventions, there are clearly unique factors to consider in the implementation of digital mental health programs.

The current study explored the perspectives of Australian secondary school staff who are primarily responsible for the adoption and delivery of such programs in the school environment. Integrating the perspectives of principals, counselors, and teachers will provide a comprehensive picture of the factors that specifically affect the implementation of digital health programs. Based on the available literature [17,18,25,26], anticipated barriers included a lack of time and curriculum constraints, limited staff to support delivery, competing responsibilities, lack of support from the leadership and other staff members, lack of knowledge or skills to facilitate delivery, and lack of perceived fit with school values; anticipated facilitators included appropriate resourcing, a clear program champion, increased school support for staff delivering the program, perceived effectiveness of the program, and knowing what implementation involves. In addition, beliefs and attitudes toward prevention and using technology to deliver mental health programs (eg, effectiveness, safety concerns, inbuilt flexibility) were expected to influence perceptions about the ease, or difficulty, of implementation. This study also explored whether factors identified by school staff act as barriers or facilitators to implementation to the same extent across different school contexts (eg, types of schools, socioeconomic status [SES]). Given the similarities in the experiences of school staff across countries (eg, high workload, burnout, role ambiguity) [28-30], the results are likely to have relevance beyond the Australian context.

Methods

Design and Recruitment

This study involved online surveys of secondary school staff from Australia, including principals and deputy principals, teachers (including librarians, support teaching staff, administrators, and Year Advisors), and school counselors and psychologists. The study procedures received ethical approval from the University of New South Wales Human Research Ethics Committee (HC17468) and the State Education Research Applications Process (SERAP2017339). Recruitment involved convenience and snowball sampling at key times (eg, beginning of school terms) between November 2017 and July 2018. Flyers and emails were sent to the research team's network of school contacts throughout New South Wales, and online flyers were posted on the Black Dog Institute's media channels (eg, Twitter, Facebook, and the Black Dog Institute website). This was an "open survey," and eligible participants were also encouraged to distribute the research opportunity directly to others in their networks.

Interested participants clicked the online link to access the survey that matched their role and were then presented with an information statement. The statement included the names of the investigators, eligibility criteria, purpose of the study, participation requirements (eg, length and nature of the survey), reimbursement, and data storage processes. Before accessing the survey, the participants provided informed consent by ticking boxes corresponding to items such as having read and understood the information statement.

The surveys were completed between 2017 and 2018 alongside recruitment. Participants had the option of providing an email address to receive reimbursement for their time with a $20AUD electronic gift card. To ensure participant anonymity and confidentiality, email addresses were obtained using a second survey accessed via a separate URL link that was provided after completing the first survey. Responses to the second survey (ie, email addresses) could not be linked to the participants' responses in the main survey.

Development and Pretesting of Measures

All the three stakeholder surveys (principal, teacher, counselor) included closed and open-ended questions about their demographics, current and previous employment, and current school. Respondents were also asked to rate how responsible they thought schools should be in addressing issues related to physical health, sexual health, mental health, and substance use among students on a Likert scale as follows: 1, very responsible; 2, somewhat responsible; 3, not really responsible; and 4, not at all responsible. A hypothetical digital depression prevention program for high school students was then described to the participants to test their attitudes toward this program. The program was described as a skill-based game, drawing from cognitive behavior therapy, that students would play over 7 20-minute sessions during class time (supervised by school staff). With all the material contained in the program, participants were informed that training or specific mental health knowledge would not be necessary. The description also emphasized that the program was evidence-based and facilitated the reduction of depressive symptoms in secondary school students. A list of barriers and facilitators to the implementation of a program of this nature were then presented. Participants were asked to rate how "challenging" the potential barriers would be for introducing the program at their school on a Likert scale as follows: 1, not at all; 2, not really; 3, moderately; 4, very; and 5, extremely. Using the same scale, they were asked to indicate how "helpful or beneficial" the potential facilitators would be for introducing this program.

For each stakeholder group, the questions were tailored to ensure suitability for their role within the school. To obtain information for developing the surveys, a review of the implementation literature pertaining to evidence-based mental health programs in schools and other relevant settings was carried out. We also consulted a network of professionals within the education, mental health, and research sectors, seeking their input regarding potential barriers and facilitators to implementing mental health programs in schools. Based on all these sources of information, we generated an extensive list of potential barriers/facilitators, which we then grouped into relevant categories, such as logistics, school support, or suitability of the program. Drafts of the surveys were circulated to our network of relevant professionals for feedback, which led to further refinement of the surveys. Once the surveys were presented electronically, they were distributed to several of the research team's contacts within each stakeholder group for testing of the usability, technical functionality, and length before recruitment for the study commenced.

Survey Administration and Procedure

The surveys were administered via Qualtrics, an online survey platform [31]. The surveys were 20-25 pages long with approximately 5-10 questions displayed per page. All questions required a response to progress to the next page. Participants could go back to review/edit responses on earlier pages without losing information already entered. Participants were also able to save their progress and continue completing the survey during subsequent sessions. A progress bar at the bottom of each page indicated approximately how far through the survey participants were, on a visual scale from 0 to 100%. Cookies were embedded on the first page of the survey, which prevented participants from using the same device to complete the survey more than once. The surveys took approximately 30-45 minutes to complete. All the survey responses were anonymous.

Statistical Analysis

All analyses were conducted using SPSS (version 25, IBM Corporation) or Excel (Microsoft Corporation). All the available participant data were analyzed, including incomplete surveys. The survey completion rate was calculated for each stakeholder group by dividing the number of people who completed the entire survey by the number of people who consented to participate. Completion rates differed between the 3 stakeholder groups, with 55% (6/11) of principals, 83% (77/93) of counselors, and 75% (73/97) of teachers completing the survey.

We calculated the descriptive statistics for each participant sample and cumulative frequencies of the perceived barriers and facilitators. We report the cumulative frequencies for barriers and facilitators rated as “moderate,” “very,” or “extreme.” This decision was taken to focus on the most frequently endorsed, and therefore practically relevant, barriers and facilitators from the perspective of key school stakeholders.

In an exploratory set of analyses, we conducted correlations to identify the relationship between key individual characteristics and beliefs about the helpfulness of online mental health programs in general. We also examined the relationship between individual characteristics and the willingness of counselors to try different mental health interventions developed by researchers. These questions were designed to assess perceptions about digital mental health programs in general. We conducted chi-square tests of independence, or the Fischer exact test of independence when expected numbers were <5, to determine whether the selected school variables were associated with perceived facilitators and barriers. The variables included the type of school (government or nongovernment), school location (capital city or rural/regional), and SES (low, medium, high). These variables are related to the availability of mental health staff and other resources in schools and can therefore affect implementation processes [32-34]. For all the analyses, the level of significance was set at $P<.05$ and strength of the association was represented by Cramer’s V.

Results

Participant Characteristics

The survey was taken by 97 teachers, 93 counselors, and 11 principals. On average, the teachers were 38.3 years old and had been in the role for 9.8 years. The school counselors were aged 39.5 years, with 6.9 years of experience in their role, and the principals were 50.9 years old, with 10 years of experience in their role. Most teachers (85/96, 89%) and counselors (62/91, 68%) were employed in government schools, whereas less than half (5/11, 46%) of the principals were from government schools. Most participants were employed fulltime (teachers: 78/97, 80%; counselors: 65/93, 70%; principals: 11/11, 100%). The perceived SES and location of employment (eg, capital city vs. rural/regional locations) varied across respondents. Table 1 presents the details.

Table 1. Summary of participant and school characteristics (given that incomplete and complete surveys were analyzed, the sample size for each characteristic may vary).

Characteristics	Teachers (n=97)	School counselors (n=93)	Principals (n=11)
Gender, female, n (%)	73 (75.3)	83 (89)	4 (36)
Age (years), mean (SD)	38.3 (11.0)	39.5 (11.2)	50.9 (7.9)
Aboriginal or Torres Strait Islander, n (%)	6 (6)	2 (2.2)	0 (0)
Born in Australia, (%)	89 (92)	79 (85)	9 (82)
Marital status, n (%)			
Married/partnered	72 (74)	82 (88)	10 (91)
Educational history, n (%)			
Postgraduate degree	41 (42)	82 (88)	9 (82)
Employment status, n (%)			
Full time	78 (80)	65 (70)	11 (100)
Years in role, mean (SD)	9.8 (9.0)	6.9 (6.5)	10.05 (12.8)
School type and location, n (%)			
Government schools	85 (89)[a]	62 (68)[c]	5 (46)
Coeducational	91 (95)[a]	75 (82)[c]	10 (91)
Low socioeconomic status	40 (42)[a]	33 (36)[c]	3 (27)
Located in a capital city	23 (24)[a]	56 (62)[c]	8 (72)
Located in a rural/regional town	73 (76)[a]	35 (39)[c]	3 (27)
Low socioeconomic status	40 (42)[b]	33 (36)[c]	3 (27)
School size, mean (SD)	767.5 (318.3)[a]	908.5 (328.9)[d]	904.2 (339.0)

[a] N=96 (for teachers).

[b] N=95 (for teachers).

[c] N=91 (for school counselors).

[d] N=89 (for school counselors).

Perceived Role of the School in Student Mental Health

All the school principals (11/11, 100%) and almost all the teachers (94/95, 99%) and counselors (89/91, 98%) felt they were either "very" or "somewhat" responsible for addressing the mental health of students. Notably, more counselors and teachers felt responsible for student mental health as compared to the number of those who felt responsible for the physical or sexual health of students, or their drug and alcohol use.

Implementation of a Digital Depression Prevention Program

For the teachers and counselors, barriers and facilitators were divided into logistical factors, school support factors, suitability of the program, and other miscellaneous factors. For the principals, to match their roles, logistical factors were conceptualized in terms of costs and resources.

Perceived Barriers

Tables S1 and S2 in Multimedia Appendix 1 display the facilitators to the implementation of a digital mental health program in schools perceived by teachers, counselors, and principals.

Logistical Factors (Including Costs and Resources)

Time was the most frequently reported barrier to program implementation by the surveyed staff, with 77% (70/91) of the teachers, 94% (84/89) of the counselors, and 64% (7/11) of the principals indicating that finding time in the school schedule would be a significant barrier. Frequencies indicated that a greater number of the surveyed counselors relative to the surveyed teachers reported that time, staff, and room availabilities were barriers to implementation. Less than half of the principals (5/11, 45%) indicated that funding for the program would be a major barrier.

School Support

Less than half of the surveyed staff members felt that obtaining school support would be a significant barrier to implementation. However, there were two exceptions for the counselors, who rated obtaining administrative support (54/89, 61%) and teacher support (53/89, 60%) as at least "moderate" challenges.

Program Suitability

Very few members of the surveyed staff were concerned that the program would increase the risk of mental illness (15% [13/88] of the teachers and 7% [6/87] of the counselors), and fewer than one-third of the teachers and counselors felt that mental health interventions should be conducted face to face rather than online. In comparison, 36% (4/11) of the principals indicated concern about the delivery format. Less than 11% of the teachers, counselors and principals felt that mental health programs were only appropriate to those with symptoms, suggesting acceptance of universal preventive approaches. A noteworthy proportion of the teachers (28/88, 32%) and counselors (40/87, 46%) expressed concerns that student privacy would be a significant barrier to the implementation of the digital program.

Other Factors

Student engagement was reported by more than half of the teachers and counselors as a potentially significant barrier to implementation. Only a small number of teachers felt that delivering this program was not their role (20/87, 23%), but many more (39/87, 45%) expressed concerns that the program would uncover mental health problems that they were not equipped to address. In contrast to the teachers, over a third of the surveyed counselors reported that delivering the program was not their role (32/87, 37%). A similar proportion of the counselors (28/87, 32%) indicated that dealing with student mental health issues and high-risk students identified during the program would be a significant barrier. Compatibility between the program and school values was also identified as a barrier by 27% (3/11) of the surveyed principals.

Perceived Facilitators

Tables S3 and S4 in Multimedia Appendix 1 display the facilitators to the implementation of a digital mental health program in schools perceived by teachers, counselors, and principals.

Logistical Factors (Including Costs and Resources)

Just over 80% of the surveyed teachers (72/86) and counselors (73/87) reported that allowing students to use personal devices to access the mental health prevention program would be helpful or beneficial to implementation. Having the program available at no cost was also rated as an important facilitator by all surveyed teachers (82/82, 100%) and most counselors (83/84, 99%), but only by 70% (7/10) of the principals.

School Support

The surveyed teachers and counselors indicated that all aspects of school support were important facilitators to program implementation. Although results were generally comparable between the teachers and counselors, having support from the principal (83/87, 95%), teachers (84/87, 97%), administrative staff (83/86, 97%), and parents (83/86, 97%) was particularly important for counselors, as was having a staff member responsible for answering concerns (83/86, 97%) and sharing responsibility for implementation with other staff members (84/86, 98%). In contrast, the surveyed principals indicated that support from counselors/well-being staff was most important (9/10, 90%).

Program Suitability

All or almost all the surveyed staff indicated that knowledge on program efficacy (for academic and emotional outcomes) would be important for implementation in schools.

Flexibility

All the surveyed principals (10/10, 100%) and most teachers (80/82, 98%) and counselors (81/84, 96%) reported that selecting an appropriate time in the school year for program delivery would be an important facilitator. Less than half of teachers (39/82, 48%) and 67% (56/84) of the counselors reported that delivering the program to only those students who might need it would be an important facilitator. Few teachers (18/82, 22%) and counselors (23/84, 27%) indicated that a targeted approach would be “very” or “extremely” important, indicating that universal approaches may be preferred.

Other Factors

The available training and the format of that training were rated by teachers (76/82, 93%) and counselors (77/84, 92%) as potential facilitators to delivery. Most surveyed teachers (76/82, 93%) and counselors (76/84, 90%) and all principals (10/10, 100%) indicated that having a screening component to identify students at risk, with student information being transferred to counselors, would be at least moderately helpful. Most surveyed teachers (72/82, 88%) indicated that aligning the program with the existing school curriculum was important, and this was comparable to the ratings provided by approximately 90% (76/84) of the counselors and principals (9/90). Alignment with the school’s philosophy, knowledge on the efficacy and benefits, and receiving feedback about the impact/success of the program from students were rated as important facilitators by all the groups.

Exploratory Analyses

Correlation analyses showed that for the surveyed teachers, age (r=0.03, P=.79) and years of experience (r=0.02, P=.88) were not associated with beliefs about how helpful online mental health programs could be in addressing common mental health problems in young people. A similar pattern of results was found for counselors (r=–0.06, P=.57 and r=–0.05, P=.67, respectively). Additionally, for counselors, there was no association between age (r=–0.16, P=.13) and years of experience (r=0.007, P=.95) with the reported willingness to try new and different types of therapy or interventions developed by researchers.

Tests of independence showed that the ratings provided for the perceived barriers and facilitators by the surveyed teachers and counselors were generally consistent across different school characteristics. However, there were some exceptions (see Tables 2-4 for descriptive percentages).

Table 2. Percentages for significant subgroup analyses with socioeconomic status.

Surveyed groups	Socioeconomic status		
	Low, n (%)	Average, n (%)	High, n (%)
Teachers			
Perceived barriers			
Time	28 (40)	34 (49)	8 (11)
Computer availability	22 (43)	26 (51)	3 (6)
Counselors			
Perceived barriers			
Obtaining administrative support	19 (35)	22 (41)	13 (24)
Perceived facilitators			
Having support from the school principal	32 (39)	27 (33)	24 (29)
Having the delivery of the program recognized by the principal as part of the job	28 (38)	25 (34)	20 (27)
Sharing responsibility for implementing the program with other staff members	31 (37)	29 (35)	24 (29)

Table 3. Percentages for significant subgroup analyses with school location.

Surveyed groups	School location	
	Capital city, n (%)	Rural/regional areas, n (%)
Teachers		
Perceived barriers		
Computer availability	14 (28)	37 (73)
Technical infrastructure and support	8 (28)	21 (72)
Counselors		
Perceived facilitators		
Having support from other teachers	50 (60)	34 (41)

Table 4. Percentages for significant subgroup analyses with school type.

Surveyed groups	School type	
	Government, n (%)	Nongovernment, n (%)
Teachers		
Perceived barriers		
Obtaining support from other teachers	68 (87.2)	2 (12.8)
Counselors		
Perceived facilitators		
Having the program aligned to the PDHPE[a] curriculum	53 (70)	23 (30)
Allowing students to use of personal devices for the program	47 (64)	26 (36)
Having the program available at no cost	66 (55)	34 (28)
Sharing responsibility for implementing the program with other staff members	57 (68)	27 (32)
Having a training manual to support implementation	55 (68)	26 (32)

[a] PDHPE: personal development, health, and physical education.

Teachers

For teachers, logistic and school support varied by subgroup. SES was significantly associated with teachers' perceptions that finding time in the school curriculum (P=.001, ν=0.33) and computer availability (P=.004, ν=0.37) would be barriers. Analyses of the percentages showed that time and computer availability were perceived as more important by teachers from low and average SES schools compared to high SES schools. The school location was significantly associated with teachers' perceptions that computer availability (*P*=.001, ν=0.55) and access to technical infrastructure (*P*=.02, ν=0.52) would be barriers. These factors were more likely to be barriers for teachers from rural/regional schools compared to schools in capital cities. The type of school was significantly associated with teachers' perceptions that obtaining support from other teachers would be a barrier (*P*=.01, ν=0.56), with teacher support being as more important in government schools than in nongovernment schools.

Counselors

The school SES was significantly associated with the counselors' perceptions that obtaining support from school administrations would be a barrier (*P*=.02, ν=0.31). Inspection of the percentages indicated that administrative support was perceived as a greater barrier by counselors from low and average SES schools compared to those in high SES schools.

In terms of facilitators, the school SES was significantly associated with the counselors' perceptions that obtaining principal support (*P*=.01, ν=0.28), recognition from the principal that program delivery was part of their job (*P*=.04, ν=0.25), and sharing responsibility for introducing the program with other staff (*P*=.04, ν=0.26) would be important facilitators. These factors were perceived as more important facilitators by counselors from low and average SES schools compared to those in high SES schools. The type of school was significantly associated with counselors' perceptions that aligning the program to the school curriculum (χ^2_2=6.27 [N=76]; *P*=.046; ν=0.29) and being able to complete the program on the students' own devices (χ^2_2=7.57 [N=73]; *P*=.02; ν=0.32) would be important facilitators. For both facilitators, there was a greater importance placed on these factors by counselors in government schools compared to those in nongovernment schools. The same pattern of results was also found for counselors' perceptions about the program being accessible for free, being provided with a training manual, and sharing responsibility for introducing the program, with *Ps*<.04 and νs>.29. Finally, the school location was significantly associated with counselors' perceptions that having support from other teachers would be an important facilitator, with *P*=.02 and ν=0.30, with teacher support being rated as a more important facilitator by counselors from schools in capital cities than those in rural/regional schools.

Discussion

Summary of Findings

The current study identified barriers and facilitators that must be considered when implementing a digital depression prevention program within Australian schools. The results align with the emergence of conceptual frameworks that outline the influential factors in the implementation of school-based interventions [35]. The results also replicate and extend findings from studies investigating the implementation and use of digital mental health programs in young populations [14,36], as well as mental health programs delivered in schools via traditional methods (eg, face-to-face programs) [17,25,37]. We clarify our results in the context of this prior research in the sections below.

Barriers to Implementation

Overall, teachers, counselors, and principals thought that a digital depression prevention program would be suitable for use within their schools, although certain factors would need to be considered to enable effective implementation. Lack of time and resources (ie, staffing and rooms) were identified as logistical barriers. This is consistent with prior work investigating the barriers for implementing face-to-face mental health programs in schools [17,25,37] and is not particularly surprising given the high workload and demands of school staff. Although endorsed by all the surveyed staff, a novel finding in our study was that these logistical barriers were particularly evident for counselors. Counselors are expected to be more familiar with what is needed to implement mental health programs effectively in schools owing to their experience in their role. Identifying differences in terms of the degree rather than the kind is an important addition to literature, with the implication being that some strategies to boost implementation ought to be tailored to specific staff members.

Teachers and counselors are concerned about information privacy in school-based digital mental health programs. Common concerns about privacy in digital approaches include how personal information about students will be stored, accessed, and used [38]. Some counselors were also concerned about having to address student mental health issues identified through the program (particularly those that are highly risky). These perceived barriers might reflect the uncertainty about how counselors will be ethically informed about students' needs and how they will provide effective care. These issues of privacy require novel solutions that are not currently addressed in face-to-face mental health care programs.

Another result supplementing the current literature was that over a third of the counselors thought that delivering the program was not their role. One explanation is that these counselors had reservations about delivering programs via digital means, and perhaps felt unequipped in terms of technical knowledge. This is an unlikely reason because most counselors (indeed, most respondents) indicated that a digital format was appropriate for students in their schools. Another possible explanation is that because the digital program would deliver standardized mental health content, counselors recognized that delivery would require low levels of support and mental health expertise, which could potentially be delegated to other school staff members such as teachers. This would benefit them in that they can have more time to work with higher-risk students and provide follow-up care. This interpretation converges with our finding that not having support from other administrative staff and teachers would be a major barrier to implementation for counselors. However, some teachers were concerned about having to help students with problems that they were not equipped to address. These results build on previous findings by indicating that new ways of collaborating with other staff might be necessary for successful implementation. Clear implementation guidelines need to be established before digital prevention programs are implemented in schools.

Facilitators to Implementation

Specific to digital programs, teachers and counselors thought that using personal devices was an important facilitator. Allowing students to use their own devices may overcome limitations in terms of school infrastructure/resources. Teachers and counselors also stated that receiving training and a training manual would assist with implementation. Training could increase staff confidence and buy-in to the program by clarifying the rationale and evidence for implementing the program, staff responsibilities, and prerequisite mental health knowledge. Previous research has shown that mental health literacy is an unmet need in school staff, with many wanting additional training to support their professional development [39-41]. Further, we discovered evidence for embedding digital programs into the national curriculum, with many school staff members indicating that existing school subjects (such as health and physical education) could be used as the timeslots for delivery. Integration into the curriculum would not only save time but also help consolidate support from the leadership (education departments, principals) and other staff within schools (eg, administration). Finally, the principals indicated that support from the counselor or well-being staff was the most important facilitator for successful implementation. Having support from staff with mental health expertise will help ensure that suitable programs are selected and that benefits are maximized for students.

Specific characteristics of the program were also identified by the surveyed staff as important facilitators. These characteristics included having an evidence-based, universal delivery, inclusion of a screening component, and clear referral pathways. A large proportion of the teachers, counselors, and principals indicated that knowledge on the program's efficacy would be important for implementation. Very few evidence-based mental health programs are implemented in schools; however, when used, they are not typically implemented as intended [34,42]. Respondents also thought that universal delivery of the program to all students, regardless of their symptoms, would increase the ease of implementation. This approach aligns with meta-analytic data showing that universal prevention programs, including those delivered using technology, are effective when delivered in schools [7].

The staff reported that incorporating a screening component to identify students at risk, with student information then being transferred to a counselor, would be necessary in the school context. Mental health screening in schools has the advantage of identifying at-risk students early and preventing them from slipping through the cracks [43-45]. In turn, engaging a triaging system would help to allocate school resources based on student needs. Previous research in Australian schools has shown that using an online mental health screening service in schools is an effective way of identifying students in need and providing an appropriate level of care [46].

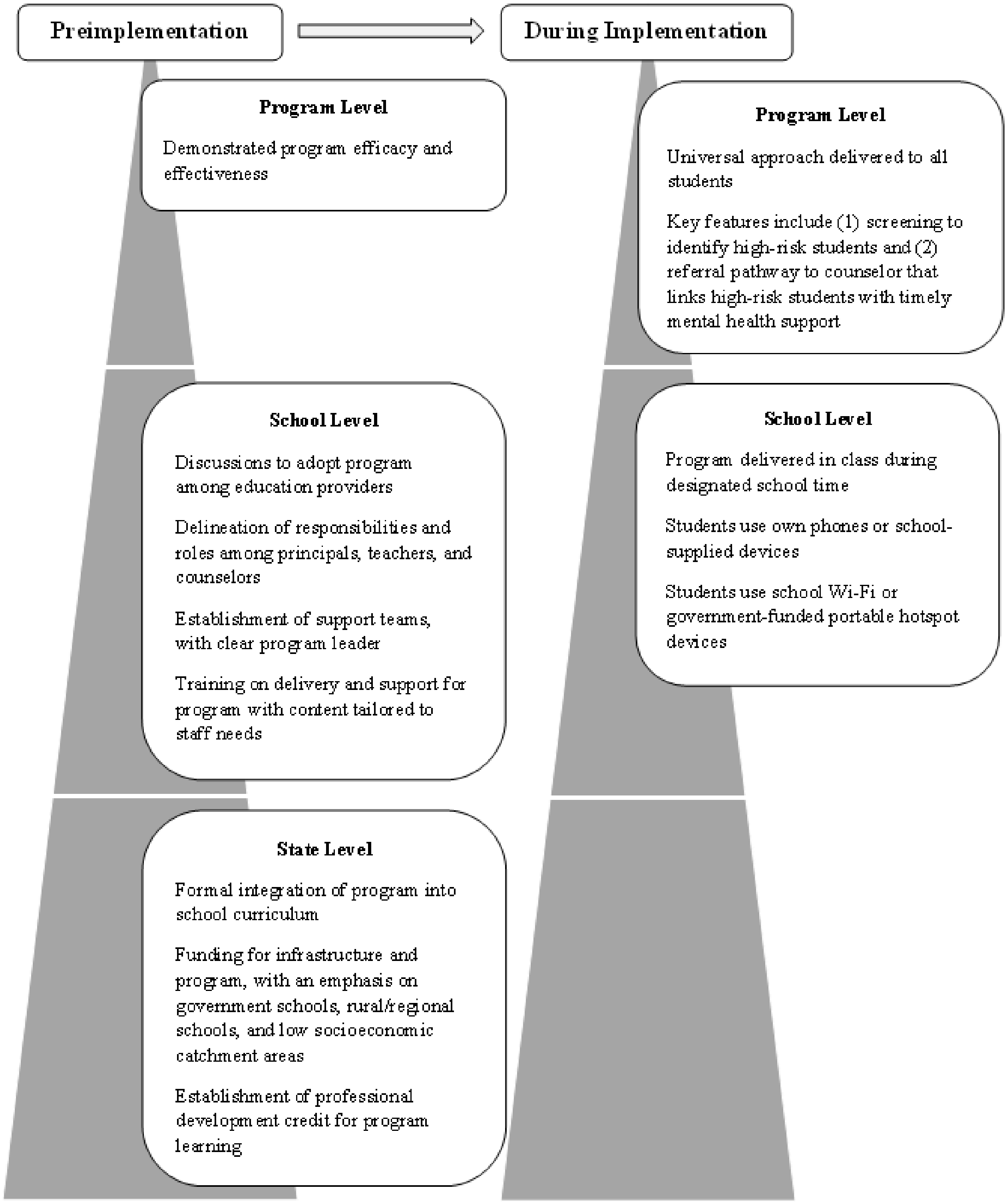
School-Based Factors

There were variations in the perceived implementation barriers and facilitators based on school contextual factors. Similar to previous research [33], the school location, type, and SES played a role in shaping the teachers' and counselors' perceptions regarding which factors would influence the implementation success. The general pattern was that the staff working in government schools, rural/regional areas, and low socioeconomic catchments anticipated facing more logistical and support barriers during implementation. These staff members identified a complementary set of facilitators to maximize implementation in these schools. Emphasis was placed on obtaining support from the other staff members and leadership, sharing responsibility for the program, embedding the program into the personal development, health, and physical education (PDHPE) curriculum, providing the program at no cost, and allowing students to use their own devices. These facilitators can be viewed as ways to reduce the burden on the already limited resources and fit the program within schools.

Implications

Our results indicate that different approaches are needed to target various staff members in different school types. One example is involving counselors in discussions about the program, before implementation, to ensure that they are on-board and can provide support. This might be particularly important in government schools. Another example is tailoring program implementation strategies to school regions and ensuring preparedness in terms of available technology, infrastructure, and support [32]. Our results also highlight the importance of factors that impact decisions related to adopting digital programs ahead of time [47]. Some face-to-face programs are already being developed with school-based barriers and facilitators in mind [48]. Extending this research to digitally delivered programs may help to design and develop school appropriate products. Figure 1 shows the proposed outline of the key implementation factors formulated from the current results.

Figure 1. Key implementation factors to consider in digital mental health program delivery at the program, school, and state levels, with focus on the Australian context.



Preimplementation

Program Level

Demonstrated program efficacy and effectiveness

School Level

Discussions to adopt program among education providers

Delineation of responsibilities and roles among principals, teachers, and counselors

Establishment of support teams, with clear program leader

Training on delivery and support for program with content tailored to staff needs

State Level

Formal integration of program into school curriculum

Funding for infrastructure and program, with an emphasis on government schools, rural/regional schools, and low socioeconomic catchment areas

Establishment of professional development credit for program learning

During Implementation

Program Level

Universal approach delivered to all students

Key features include (1) screening to identify high-risk students and (2) referral pathway to counselor that links high-risk students with timely mental health support

School Level

Program delivered in class during designated school time

Students use own phones or school-supplied devices

Students use school Wi-Fi or government-funded portable hotspot devices

Limitations

This study is limited in some respects. The sampling processes could have contributed to selection biases and positive self-presentation. School staff who responded to the survey may have already been interested or engaged in the mental health of their students. Staff members who were reached by recruitment but did not consent to participate may have had systematically different perspectives about the anticipated barriers and facilitators. Further, our study only included a small proportion of the total number of staff members in Australia eligible to participate in the study. However, the teachers, counselors, and principals in our study did represent a variety of school types, locations, and socioeconomic circumstances, which increases the generalizability of our findings. The characteristics of our sample, including the average age and gender proportion, were also consistent with the population characteristics of Australian teachers and principals [49]. Random sampling with more participants will help to prevent biases in sampling during future studies.

Another limitation is that we only focused on schools in Australia. Although our results may be relevant in other Australian states and countries with similar school systems (eg, United Kingdom), it is unclear how digital programs would be received in places with fewer resources or different staffing roles. Future research could compare the implementation factors in schools on an international level, with particular focus on schools in low- and middle-income countries (LMICs). LMICs are likely to encounter different barriers and facilitators, issues that are important to consider when implementing such programs on a large scale [50]. Finally, given that we examined a hypothetical digital depression prevention program, our conclusions may not be generalizable to the actual implementation of real-world programs. Implementation process evaluations of digital mental health programs will help understand what the most effective approach in schools is, thereby improving outcomes for students and school communities. One example of such an evaluation is currently underway in Australian schools [16,51].

Conclusions

Our study explored the factors influencing the implementation of a digital depression prevention program within secondary schools from the perspectives of teachers, counselors, and principals. The surveyed staff members thought that a digital program focusing on universal prevention was suitable for delivery in their schools. A range of logistical, support, and program barriers and facilitators were identified, some of which were unique to digital programs. There were also some differences between what teachers, counselors, and principals thought were the most important factors to be considered, as well as among staff members from different schools. Our results highlight the importance of considering multiple perspectives in school-based implementation and tailoring strategies to maximize digital delivery based on staff roles and school types.

Acknowledgments

AW-S secured funding from a New South Wales Health Early-Mid Career Fellowship. The funding body had no role in any aspect of the study design or in writing this manuscript. We would also like to thank Katarina Kikas, who assisted with the development of Figure 1.

Authors' Contributions

AW-S conceived the study. LJ collected the data. JRB conducted the data analysis and wrote the manuscript, with assistance from AW-S. All authors contributed to the methodology, including the development of the survey measures. All authors have read, reviewed, refined, and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Result tables for barriers and facilitators.
[PDF File (Adobe PDF File), 161 KB - jmir_v23i8e26223_app1.pdf]

References

1. Merikangas KR, He JP, Brody D, Fisher PW, Bourdon K, Koretz DS. Prevalence and treatment of mental disorders among US children in the 2001-2004 NHANES. Pediatrics 2010 Jan;125(1):75-81 [FREE Full text] [doi: 10.1542/peds.2008-2598] [Medline: 20008426]
2. Dunn V, Goodyer IM. Longitudinal investigation into childhood- and adolescence-onset depression: psychiatric outcome in early adulthood. Br J Psychiatry 2006 Mar;188(3):216-222. [doi: 10.1192/bjp.188.3.216] [Medline: 16507961]
3. Ormel J, Kessler R, Schoevers R. Depression: more treatment but no drop in prevalence: How effective is treatment? And can we do better? Curr Opin Psychiatry 2019 Jul;32(4):348-354. [doi: 10.1097/YCO.0000000000000505] [Medline: 30855297]
4. Farmer EMZ, Burns BJ, Phillips SD, Angold A, Costello EJ. Pathways into and through mental health services for children and adolescents. Psychiatr Serv 2003 Jan;54(1):60-66. [doi: 10.1176/appi.ps.54.1.60] [Medline: 12509668]
5. Burns BJ, Costello EJ, Angold A, Tweed D, Stangl D, Farmer EM, et al. Children's mental health service use across service sectors. Health Aff (Millwood) 1995 Jan;14(3):147-159. [doi: 10.1377/hlthaff.14.3.147] [Medline: 7498888]
6. Stockings EA, Degenhardt L, Dobbins T, Lee YY, Erskine HE, Whiteford HA, et al. Preventing depression and anxiety in young people: a review of the joint efficacy of universal, selective and indicated prevention. Psychol Med 2015 Aug;46(1):11-26. [doi: 10.1017/s0033291715001725]
7. Werner-Seidler A, Perry Y, Calear AL, Newby JM, Christensen H. School-based depression and anxiety prevention programs for young people: a systematic review and meta-analysis. Clin Psychol Rev 2017 Feb;51:30-47 [FREE Full text] [doi: 10.1016/j.cpr.2016.10.005] [Medline: 27821267]

8. Calear AL, Christensen H, Mackinnon A, Griffiths KM, O'Kearney R. The YouthMood project: a cluster randomized controlled trial of an online cognitive behavioral program with adolescents. J Consult Clin Psychol 2009 Dec;77(6):1021-1032. [doi: 10.1037/a0017391] [Medline: 19968379]
9. Wong N, Kady L, Mewton L, Sunderland M, Andrews G. Preventing anxiety and depression in adolescents: a randomised controlled trial of two school based Internet-delivered cognitive behavioural therapy programmes. Internet Interv 2014 Apr;1(2):90-94. [doi: 10.1016/j.invent.2014.05.004]
10. David OA, Cardoş RAI, Matu S. Is REThink therapeutic game effective in preventing emotional disorders in children and adolescents? outcomes of a randomized clinical trial. Eur Child Adolesc Psychiatry 2019 Jan;28(1):111-122. [doi: 10.1007/s00787-018-1192-2] [Medline: 29992353]
11. Perry Y, Werner-Seidler A, Calear A, Mackinnon A, King C, Scott J, et al. Preventing depression in final year secondary students: school-based randomized controlled trial. J Med Internet Res 2017 Nov;19(11):e369 [FREE Full text] [doi: 10.2196/jmir.8241] [Medline: 29097357]
12. Poppelaars M, Tak YR, Lichtwarck-Aschoff A, Engels RC, Lobel A, Merry SN, et al. A randomized controlled trial comparing two cognitive-behavioral programs for adolescent girls with subclinical depression: a school-based program (Op Volle Kracht) and a computerized program (SPARX). Behav Res Ther 2016 May;80:33-42. [doi: 10.1016/j.brat.2016.03.005] [Medline: 27019280]
13. Whittaker R, Stasiak K, McDowell H, Doherty I, Shepherd M, Chua S, et al. MEMO: An mHealth intervention to prevent the onset of depression in adolescents: a double-blind, randomised, placebo-controlled trial. J Child Psychol Psychiatry 2017 Sep 02;58(9):1014-1022. [doi: 10.1111/jcpp.12753] [Medline: 28573672]
14. Liverpool S, Mota CP, Sales CMD, Čuš A, Carletto S, Hancheva C, et al. Engaging children and young people in digital mental health interventions: systematic review of modes of delivery, facilitators, and barriers. J Med Internet Res 2020 Jun;22(6):e16317 [FREE Full text] [doi: 10.2196/16317] [Medline: 32442160]
15. Ellis LA, Augustsson H, Grødahl AI, Pomare C, Churruca K, Long JC, et al. Implementation of e-mental health for depression and anxiety: a critical scoping review. J Community Psychol 2020 Apr;48(3):904-920. [doi: 10.1002/jcop.22309] [Medline: 31944324]
16. Beames JR, Lingam R, Boydell K, Calear AL, Torok M, Maston K, et al. Protocol for the process evaluation of a complex intervention delivered in schools to prevent adolescent depression: the future proofing study. BMJ Open 2021 Jan;11(1):e042133 [FREE Full text] [doi: 10.1136/bmjopen-2020-042133] [Medline: 33436468]
17. Forman SG, Olin SS, Hoagwood KE, Crowe M, Saka N. Evidence-based interventions in schools: developers’ views of implementation barriers and facilitators. Sch Ment Health 2008 Nov;1(1):26-36. [doi: 10.1007/s12310-008-9002-5]
18. Langley AK, Nadeem E, Kataoka SH, Stein BD, Jaycox LH. Evidence-based mental health programs in schools: barriers and facilitators of successful implementation. School Ment Health 2010 Sep;2(3):105-113 [FREE Full text] [doi: 10.1007/s12310-010-9038-1] [Medline: 20694034]
19. Patalay P, Giese L, Stanković M, Curtin C, Moltrecht B, Gondek D. Mental health provision in schools: priority, facilitators and barriers in 10 European countries. Child Adolesc Ment Health 2016 Sep;21(3):139-147. [doi: 10.1111/camh.12160] [Medline: 32680356]
20. Dariotis JK, Mirabal-Beltran R, Cluxton-Keller F, Gould LF, Greenberg MT, Mendelson T. A qualitative exploration of implementation factors in a school-based mindfulness and yoga program: lessons learned from students and teachers. Psychol Sch 2017 Jan;54(1):53-69 [FREE Full text] [doi: 10.1002/pits.21979] [Medline: 28670007]
21. Lyon AR, Charlesworth-Attie S, Vander Stoep A, McCauley E. Modular psychotherapy for youth with internalizing problems: implementation with therapists in school-based health centers. Sch Psychol Rev 2019 Dec;40(4):569-581. [doi: 10.1080/02796015.2011.12087530]
22. Mendelson T, Dariotis JK, Feagans Gould L, Smith ASR, Smith AA, Gonzalez AA, et al. Implementing mindfulness and yoga in urban schools: a community-academic partnership. J Child Serv 2013 Nov;8(4):276-291. [doi: 10.1108/jcs-07-2013-0024]
23. Connors EH, Arora P, Curtis L, Stephan SH. Evidence-based assessment in school mental health. Cogn Behav Pract 2015 Feb;22(1):60-73. [doi: 10.1016/j.cbpra.2014.03.008]
24. Hicks TB, Shahidullah JD, Carlson JS, Palejwala MH. Nationally certified school psychologists' use and reported barriers to using evidence-based interventions in schools: the influence of graduate program training and education. Sch Psychol Q 2014 Dec;29(4):469-487. [doi: 10.1037/spq0000059] [Medline: 24708281]
25. Hudson KG, Lawton R, Hugh-Jones S. Factors affecting the implementation of a whole school mindfulness program: a qualitative study using the consolidated framework for implementation research. BMC Health Serv Res 2020 Feb;20(1):133 [FREE Full text] [doi: 10.1186/s12913-020-4942-z] [Medline: 32087729]
26. Wilde S, Sonley A, Crane C, Ford T, Raja A, Robson J, et al. Mindfulness training in UK secondary schools: a multiple case study approach to identification of cornerstones of implementation. Mindfulness 2019 Jun;10(2):376-389 [FREE Full text] [doi: 10.1007/s12671-018-0982-4] [Medline: 31186817]
27. Short KH. Intentional, explicit, systematic: implementation and scale-up of effective practices for supporting student mental well-being in Ontario schools. Int J Ment Health Promot 2016 Jan;18(1):33-48 [FREE Full text] [doi: 10.1080/14623730.2015.1088681] [Medline: 27019639]

28. Arvidsson I, Leo U, Larsson A, Håkansson C, Persson R, Björk J. Burnout among school teachers: quantitative and qualitative results from a follow-up study in southern Sweden. BMC Public Health 2019 May;19(1):655 [FREE Full text] [doi: 10.1186/s12889-019-6972-1] [Medline: 31142318]
29. Carod-Artal F, Vázquez-Cabrera C. Burnout syndrome in an international setting. In: Bährer-Kohler S, editor. Burnout for Experts: Prevention in the Context of Living and Working. New York: Springer Science & Business Media; 2013:15-35.
30. Lawrence DF, Loi NM, Gudex BW. Understanding the relationship between work intensification and burnout in secondary teachers. Teachers and Teaching 2018 Nov;25(2):189-199. [doi: 10.1080/13540602.2018.1544551]
31. Qualtrics computer program. 2017. URL: http://www.qualtrics.com [accessed 2020-06-16]
32. Eiraldi R, Wolk CB, Locke J, Beidas R. Clearing hurdles: the challenges of implementation of mental health evidence-based practices in under-resourced schools. Adv Sch Ment Health Promot 2015 May;8(3):124-140. [doi: 10.1080/1754730x.2015.1037848]
33. Payne AA, Gottfredson DC, Gottfredson GD. School predictors of the intensity of implementation of school-based prevention programs: results from a national study. Prev Sci 2006 Jun;7(2):225-237. [doi: 10.1007/s11121-006-0029-2] [Medline: 16791524]
34. Gottfredson DC, Gottfredson GD. Quality of school-based prevention programs: results from a national survey. J Res Crime Delinq 2016 Aug;39(1):3-35. [doi: 10.1177/002242780203900101]
35. Domitrovich CE, Bradshaw CP, Poduska JM, Hoagwood K, Buckley JA, Olin S, et al. Maximizing the implementation quality of evidence-based preventive interventions in schools: a conceptual framework. Adv Sch Ment Health Promot 2008 Jul;1(3):6-28 [FREE Full text] [doi: 10.1080/1754730x.2008.9715730] [Medline: 27182282]
36. Cheek C, Bridgman H, Fleming T, Cummings E, Ellis L, Lucassen MF, et al. Views of young people in rural Australia on SPARX, a fantasy world developed for New Zealand youth with depression. JMIR Serious Games 2014 Feb;2(1):e3 [FREE Full text] [doi: 10.2196/games.3183] [Medline: 25659116]
37. Langley AK, Nadeem E, Kataoka SH, Stein BD, Jaycox LH. Evidence-based mental health programs in schools: barriers and facilitators of successful implementation. Sch Ment Health 2010 Sep 11;2(3):105-113 [FREE Full text] [doi: 10.1007/s12310-010-9038-1] [Medline: 20694034]
38. 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]
39. Frauenholtz S, Mendenhall AN, Moon J. Role of school employees' mental health knowledge in interdisciplinary collaborations to support the academic success of students experiencing mental health distress. Children Sch 2017 Feb;39(2):71-79. [doi: 10.1093/cs/cdx004]
40. Mazzer KR, Rickwood DJ. Teachers' role breadth and perceived efficacy in supporting student mental health. Adv Sch Ment Health Promot 2014 Nov;8(1):29-41. [doi: 10.1080/1754730x.2014.978119]
41. Willis A, Hyde M, Black A. Juggling with both hands tied behind my back: teachers' views and experiences of the tensions between student well-being concerns and academic performance improvement agendas. Am Educ Res J 2019 May;56(6):2644-2673. [doi: 10.3102/0002831219849877]
42. Hallfors D, Godette D. Will the 'principles of effectiveness' improve prevention practice? early findings from a diffusion study. Health Educ Res 2002 Aug;17(4):461-470. [doi: 10.1093/her/17.4.461] [Medline: 12197591]
43. Siceloff ER, Bradley WJ, Flory K. Universal behavioral/emotional health screening in schools: overview and feasibility. Rep Emot Behav Disord Youth 2017;17(2):32-38 [FREE Full text] [Medline: 30705612]
44. Weist MD, Rubin M, Moore E, Adelsheim S, Wrobel G. Mental health screening in schools. J Sch Health 2007 Feb;77(2):53-58. [doi: 10.1111/j.1746-1561.2007.00167.x] [Medline: 17222155]
45. Splett JW, Trainor KM, Raborn A, Halliday-Boykins CA, Garzona ME, Dongo MD, et al. Comparison of universal mental health screening to students already receiving intervention in a multitiered system of support. Behav Disord 2018 Apr;43(3):344-356. [doi: 10.1177/0198742918761339]
46. O'Dea B, King C, Subotic-Kerry M, Achilles MR, Cockayne N, Christensen H. Smooth sailing: a pilot study of an online, school-based, mental health service for depression and anxiety. Front Psychiatry 2019 Aug;10:574 [FREE Full text] [doi: 10.3389/fpsyt.2019.00574] [Medline: 31481904]
47. Powers JD, Bowen NK, Bowen GL. Evidence-based programs in school settings: barriers and recent advances. J Evid Based Soc Work 2010 Jul;7(4):313-331. [doi: 10.1080/15433710903256807] [Medline: 20799130]
48. Pina AA, Gonzales NA, Mazza GL, Gunn HJ, Holly LE, Stoll RD, et al. Streamlined prevention and early intervention for pediatric anxiety disorders: a randomized controlled trial. Prev Sci 2020 May;21(4):487-497 [FREE Full text] [doi: 10.1007/s11121-019-01066-6] [Medline: 31927654]
49. Thompson S, Hillman K. The Teaching and Learning International Survey 2018. Australian Report Volume 1: Teachers and School Leaders as Lifelong Learners. Camberwell, Australia: Australian council for educational research (ACER); 2019.
50. Fazel M, Patel V, Thomas S, Tol W. Mental health interventions in schools in low-income and middle-income countries. The Lancet Psychiatry 2014 Oct;1(5):388-398. [doi: 10.1016/s2215-0366(14)70357-8]

51. Werner-Seidler A, Huckvale K, Larsen ME, Calear AL, Maston K, Johnston L, et al. A trial protocol for the effectiveness of digital interventions for preventing depression in adolescents: the future proofing study. Trials 2020 Jan;21(1):2 [FREE Full text] [doi: 10.1186/s13063-019-3901-7] [Medline: 31898512]

Abbreviations

LMIC: low- and middle-income country
PDHPE: personal development, health, and physical education
SES: socioeconomic status

Edited by R Kukafka; submitted 02.12.20; peer-reviewed by W Guo, L Birrell; comments to author 11.01.21; revised version received 17.01.21; accepted 04.05.21; published 27.08.21.

Please cite as:
Beames JR, Johnston L, O'Dea B, Torok M, Christensen H, Boydell KM, Werner-Seidler A
Factors That Help and Hinder the Implementation of Digital Depression Prevention Programs: School-Based Cross-sectional Study
J Med Internet Res 2021;23(8):e26223
URL: https://www.jmir.org/2021/8/e26223
doi:10.2196/26223
PMID:34448701

©Joanne R Beames, Lara Johnston, Bridianne O'Dea, Michelle Torok, Helen Christensen, Katherine M Boydell, Aliza Werner-Seidler. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Using Machine Learning–Based Approaches for the Detection and Classification of Human Papillomavirus Vaccine Misinformation: Infodemiology Study of Reddit Discussions

Jingcheng Du[1*], PhD; Sharice Preston[2*], PhD; Hanxiao Sun[1,2], MSc; Ross Shegog[2], PhD; Rachel Cunningham[3], MSc; Julie Boom[3,4], MD; Lara Savas[2], PhD; Muhammad Amith[1], PhD; Cui Tao[1], PhD

[1]School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, United States
[2]School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, United States
[3]Texas Children's Hospital, Houston, TX, United States
[4]Baylor College of Medicine, Houston, TX, United States
[*]these authors contributed equally

Corresponding Author:
Cui Tao, PhD
School of Biomedical Informatics
The University of Texas Health Science Center at Houston
7000 Fannin St
#600
Houston, TX, 77030
United States
Phone: 1 713 500 981
Email: cui.tao@uth.tmc.edu

Abstract

Background: The rapid growth of social media as an information channel has made it possible to quickly spread inaccurate or false vaccine information, thus creating obstacles for vaccine promotion.

Objective: The aim of this study is to develop and evaluate an intelligent automated protocol for identifying and classifying human papillomavirus (HPV) vaccine misinformation on social media using machine learning (ML)–based methods.

Methods: Reddit posts (from 2007 to 2017, N=28,121) that contained keywords related to HPV vaccination were compiled. A random subset (2200/28,121, 7.82%) was manually labeled for misinformation and served as the gold standard corpus for evaluation. A total of 5 ML-based algorithms, including a support vector machine, logistic regression, extremely randomized trees, a convolutional neural network, and a recurrent neural network designed to identify vaccine misinformation, were evaluated for identification performance. Topic modeling was applied to identify the major categories associated with HPV vaccine misinformation.

Results: A convolutional neural network model achieved the highest area under the receiver operating characteristic curve of 0.7943. Of the 28,121 Reddit posts, 7207 (25.63%) were classified as vaccine misinformation, with discussions about general safety issues identified as the leading type of misinformed posts (2666/7207, 36.99%).

Conclusions: ML-based approaches are effective in the identification and classification of HPV vaccine misinformation on Reddit and may be generalizable to other social media platforms. ML-based methods may provide the capacity and utility to meet the challenge involved in intelligent automated monitoring and classification of public health misinformation on social media platforms. The timely identification of vaccine misinformation on the internet is the first step in misinformation correction and vaccine promotion.

(J Med Internet Res 2021;23(8):e26478) doi:10.2196/26478

KEYWORDS

HPV vaccine; social media; misinformation; infodemiology; infoveillance; deep learning; Reddit; machine learning

Introduction

Background

Human papillomavirus (HPV) infection is a highly prevalent sexually transmitted infection. HPV infections cause approximately 33,700 cases of cancer every year in the United States, including cervical, vaginal, penile, anal, and head and neck cancers [1,2]. Since 2006, a vaccine against the most common HPV subtypes has been available to prevent associated cancers and genital warts [3]. Despite undeniable evidence of its effectiveness, the HPV vaccine has been controversial among parents, which has contributed to vaccine hesitancy and even refusal [4] and to relatively low national rates of HPV vaccine initiation and series completion [5]. Resistance to the HPV vaccine has been a result of parents' concerns about the vaccine's effect on sexual behavior, because HPV is a sexually transmitted infection, and the safety of the vaccine, as well as inconsistent vaccine recommendations from health care providers [6].

A burgeoning antivaccine movement has affected overall vaccine coverage in the United States and contributed to a resurgence of vaccine-preventable diseases such as measles [7]. Vaccine hesitancy has been found to be driven mainly by concerns about vaccine safety and is propelled by misinformation circulated through social media [8]. The rapid growth of social media as an information channel has made it possible to quickly spread inaccurate or false information and create a platform for antivaccine campaigns to promulgate vaccine-related misinformation [9]. Participants in the antivaccine movement circulate antivaccine sentiments and misinformation through various internet channels and create demonstrable impact on individual and community health [10]. Experts in media communications have suggested that web-based misinformation is becoming unmanageable, even as concern increases about the damage it causes to consumer well-being [11]. Efforts to curtail the phenomenon, such as story-flagging and fact-checking tools, are not enough to suppress the advocates of misinformation [12,13] because the efficiency and scalability of these tools are limited, and misinformation is disseminated much faster and broader than true information.

Mitigation of medical and public health misinformation on social media is important; however, the sheer amount of information makes it challenging to identify these posts efficiently and accurately. Although social media is a convenient way for users to generate, share, receive, and comment on social content [14], there is a need for broad-scale, innovative methods to track and understand the spread of health misinformation on social media outlets [15].

Identifying vaccine-related misinformation presented on social media is an important first step in the timely curbing of the ongoing spread of vaccine misinformation. Given the large volume of social media posts and unique features of social media language (ie, incomplete sentences and misspellings), the use of automated methods for the identification of misinformation is challenging. However, machine learning (ML)–based approaches have been previously applied to identify misinformation on Twitter regarding controversial topic domains [16] and rumors regarding a range of topics [17]. ML involves the use of algorithms and statistical modeling that provide the ability to automatically conduct tasks and learn without using explicit programming [18]. Despite the utility of these ML approaches, there is a dearth of application to medical or health topics. To date, an ML-based system has tracked misinformation about the Zika virus on social media [19] and classified misinformation within specific health forums (eg, MedHelp) [20]. Although there have been efforts to develop ML for sentiment analysis on vaccine topics [21,22], to our knowledge, there is no prior work on *automated identification* of vaccine-related misinformation on social media. Deep learning (DL) is a subset of ML algorithms based on deep neural networks. Although DL has advanced ML algorithms in multiple tasks [23], the utility of DL regarding vaccine misinformation identification is still unclear.

Objective

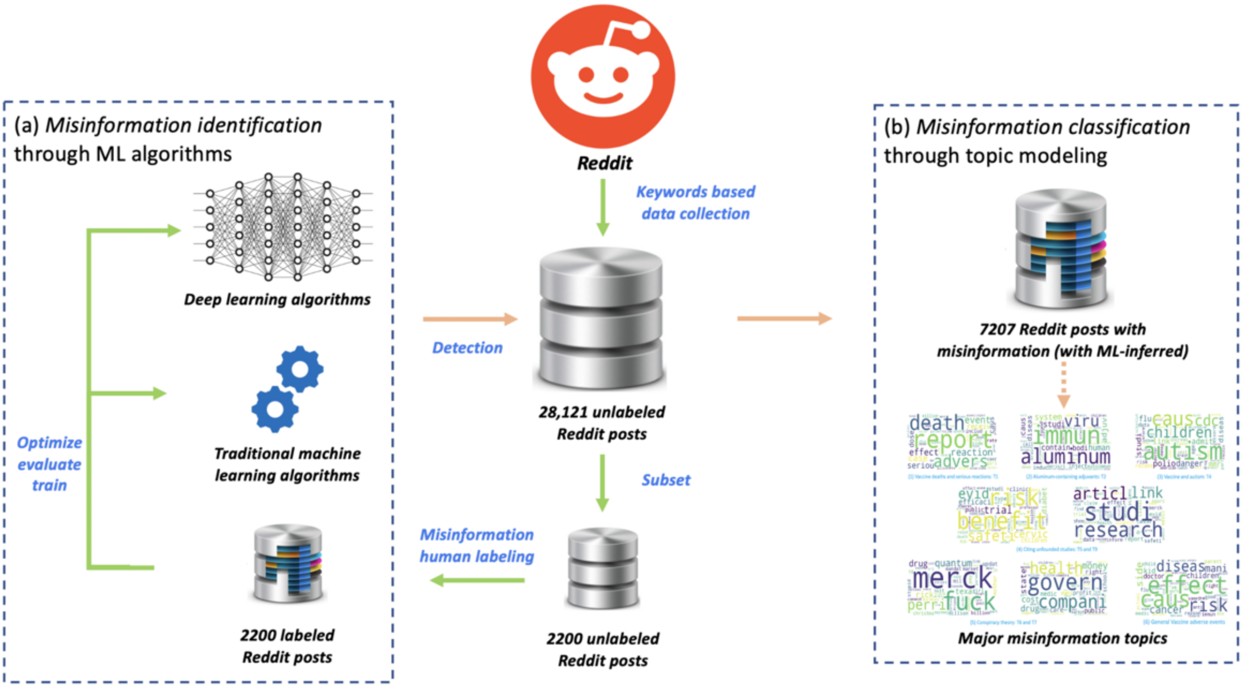
We report the utility of various conventional ML and DL algorithms to automatically identify and categorize misinformation on the HPV vaccine using posts on Reddit, a popular social media platform with more than 330 million monthly active users [24]. Reddit users are primarily anonymous young users below the age of 35 years, and more than half (54%) live in the United States [25]). Studies have revealed that young adults in the United States have low perceived susceptibility of contracting HPV [26], low health literacy pertaining to HPV and the HPV vaccine [27], and are more likely to seek health information on social media than other age demographics [28]. Table S1 in Multimedia Appendix 1 lists ML-related terms in the manuscript and their definitions.

Methods

Overview

We used a hybrid approach for the identification and classification of HPV vaccine misinformation on Reddit (Figure 1). Our approach can be divided into two steps: (1) evaluation of ML algorithms for vaccine misinformation identification and (2) topic modeling on Reddit posts that contain vaccine misinformation (ML-inferred).

Figure 1. The overview of human papillomavirus misinformation identification and classification on Reddit. (a) Evaluation of machine learning–based misinformation identification and (b) topic modeling. ML: machine learning.


(a) Misinformation identification through ML algorithms
Deep learning algorithms
Traditional machine learning algorithms
Optimize evaluate train
2200 labeled Reddit posts
Reddit
Keywords based data collection
Detection
28,121 unlabeled Reddit posts
Subset
Misinformation human labeling
2200 unlabeled Reddit posts
(b) Misinformation classification through topic modeling
7207 Reddit posts with misinformation (with ML-inferred)
Major misinformation topics

Reddit posts that contain HPV vaccine keywords were first collected, after which random subsets of the posts were labeled as containing *misinformation* or *nonmisinformation*. The labeled Reddit posts served as the gold standard corpus for the training and evaluation of various traditional ML and DL algorithms. The best algorithm, one that achieved the highest area under the receiver operating characteristic curve (AUC) [29], was then selected to infer the misinformation label for the remaining unlabeled Reddit posts. Finally, we applied topic modeling to the Reddit posts that were classified as misinformation to explore the major discussion topics and their prevalence.

Data Set Collection and Labeling

We collected Reddit discussions related to HPV vaccination from 2007 to 2017 (N=28,121) using Pushshift [30]. Submissions (topic starters) and comments (responses to the topic) that contained the case-insensitive expressions of *cervarix* or *gardasil* or the combination of *hpv* or *papillomavirus* with *shot* or *vaccine* were collected.

We selected a random sample (simple random sampling) of the total collected Reddit posts (2200/28,121, 7.82%) and labeled the posts as misinformation or nonmisinformation. The purpose of this step was to build a gold standard corpus (ie, Reddit posts with their expert-assigned labels) that was used for the training and evaluation of the automated ML algorithms. The definition of vaccine misinformation was largely informed by the *Antivaccination Information* class of the Vaccine Misinformation Ontology (VAXMO) [31], a formal ontology to describe vaccine misinformation. Within the VAXMO, the *Antivaccination Information* class includes several subclasses, such as *Vaccine inefficacy*, *Alternative medicine*, *Civil liberties*, *Conspiracy theories*, *Falsehoods*, and *Ideological*. The random sample of the Reddit posts was used to develop a guideline through discussion among the annotators. A priori consensus was reached among 3 of the study annotators to combine the subclasses *Civil Liberties* and *Ideological* and to add two categories: *Vaccine recommendations* and *Other.* The resultant decision rules were that if a Reddit post contained one or more types of vaccine misinformation, it was considered an instance of misinformation (Textbox 1).

Textbox 1. Descriptions of types of vaccine misinformation.

- Vaccine inefficacy: vaccine misinformation related to concerns about the lack of effectiveness of vaccines.
- Vaccine safety: vaccine misinformation related to concerns regarding safety issues and supposed harmful ingredients.
- Conspiracy theories: vaccine misinformation related to accusations of a cover-up, where regulatory bodies purportedly have information about vaccines that they are hiding from the public.
- Vaccine recommendations: vaccine misinformation related to vaccine recommendation or schedule.
- Civil liberties and ideologies: the encroachment on personal and parental legal rights or personal principles influencing individual opinions about antivaccine sentiment based on religion, morality, or other ideological reasons.
- Other: other types of vaccine misinformation or a mixed type of misinformation.

Furthermore, 3 study team members in the fields of biomedical informatics and public health (JD, SP, and HS) were involved in the annotation. The first 100 Reddit posts served as training, with the annotators independently annotating each post and then discussing each post and its annotation as a group. The training ended when the annotators achieved consensus (or took a decision based on a majority vote) on all the posts. After the annotation training, the remaining sampled Reddit posts were split among the 3 annotators for independent labeling. To examine the quality of the annotation, we selected 200 additional posts from the unlabeled Reddit posts, and JD, SP, and HS worked on these posts independently. We calculated the Cohen κ among the 3 annotators [32]. The total labeled Reddit samples were used as the basis of a gold standard corpus that was subsequently used for the training and evaluation of the automated ML algorithms.

Misinformation Identification

Text Classification

Text classification is a fundamental task of natural language processing (NLP) which aims to classify the textual posts into predefined classes [33]. NLP is a subfield of artificial intelligence that allows computers to process and analyze natural language (ie, free text) data. We framed the identification of misinformation from Reddit posts as a binary text classification task. Each Reddit post was assigned one of two exclusive labels (ie, misinformation or nonmisinformation) within the automated ML-based algorithms (described below).

ML Algorithms

We evaluated 5 ML-based algorithms: 3 conventional and 2 DL algorithms. Traditional ML algorithms (ie, nondeep neural network–based algorithms) with feature engineering are widely used for text classification tasks. Altogether, 3 conventional ML algorithms were evaluated in this study: a support vector machine, logistic regression (LR), and extremely randomized trees. Support vector machines have been widely used in text classification tasks [34-36]. LR has achieved favorable performance on many task classification tasks as well but requires substantially less running time [37,38]. Extremely randomized trees is a tree-based ensemble method that has achieved favorable performances in our previous studies on social media text classification tasks [39,40]. Term frequency-inverse document frequency (TF-IDF) was adopted as the feature for these traditional ML algorithms. TF-IDF is a numerical statistic that assesses the relative importance of a word to a document in a corpus [41].

DL is a subset of ML algorithms. We evaluated 2 commonly used DL-based frameworks in this study: convolutional neural network (CNN) [42] and recurrent neural network (RNN) [43]. The effectiveness of traditional ML algorithms depends on task-specific feature engineering [44]. Deep neural networks can take advantage of pretrained word embedding to capture the semantics of words, which saves significant effort in feature engineering by domain experts [45]. DL algorithms have achieved state-of-the-art performance on many text classification tasks [46-49].

As there are frequent occurrences of incorrect spelling in social media posts, both the evaluated DL algorithms contained a character layer and a word-embedding layer to map both in-vocabulary (ie, correctly spelled) and out-of-vocabulary (ie, incorrectly spelled) words to high-dimensional vectors to represent their semantics. GloVe (Global Vectors for Word Representation) embedding (ie, glove.840B.300d) [50] was used to initialize the weights in the word-embedding layer. The CNN model takes word-level embedding as input and feeds it to convolution and max-pooling layers, a fully connected layer and a softmax layer, respectively, for classification [42]. The RNN model follows an architecture that is similar to that of the CNN model by replacing convolution and max-pooling layers with bidirectional long short-term memory layers and attention layers.

More specifically, for both the CNN and RNN models, the learning rate was set at 0.01, the batch size was 64, and the number of epochs was set at 100. The length of the character embedding was set at 50 for both models. For the CNN model, the filter sizes were 1, 2, and 3, and the number of filters was 2048; for the RNN model, the hidden dimension of the long short-term memory unit and attention layer was set to 128. The dropout probability was 0.2 for both models. The model that achieved the best AUC value on the validation set was selected for testing and prediction.

Experiment Settings and Evaluation

The gold standard posts (ie, Reddit posts with expert-assigned labels) were randomly split into train, validation, and test sets in a ratio of 7:1:2. We adopted spaCy *tokenization* [51] to split the post text into separate words, remove punctuations, and convert words and letters in uppercase to lowercase. Sequentially, the train set was used to *train* the algorithms, the validation set was used for hyperparameter selection, and the test set was used to evaluate the performance of the models. To account for imbalance in label distribution, the criterion of reference was the degree of specificity as measured by the optimal AUC. The algorithm with the highest AUC value was selected for the inference of vaccine misinformation from unlabeled Reddit posts.

We further plotted the precision and recall curves of the best-performing model (ie, the CNN model) and selected the optimal cutoff (based on the highest F_1 score) of the algorithm to identify vaccine misinformation in Reddit posts. Precision was defined as the fraction of misinformation posts identified by the labelers among the fraction of misinformation posts identified by the classifier. Recall was defined as the fraction of misinformation posts identified by the labelers that were retrieved by the classifier. The F_1 score is a harmonic mean of precision and recall. The cutoff that led to the best F_1 score for the CNN model was selected. The model was applied to identify vaccine misinformation–related Reddit posts in the remaining unlabeled Reddit corpus.

Misinformation Topic Model

The ML and DL algorithms described above can potentially be used to automatically identify Reddit posts with misinformation, but they do not categorize the types of misinformation. We adopted a topic model algorithm (ie, Biterm Topic model [BTM]) [52], and we implemented the code from a GitHub repository [53]) to identify and visualize major topics from the misinformation in Reddit posts. Topic models are a type of statistical model designed to cluster the abstract *topics* that occur in a collection of documents. After using the best-performing ML algorithm to identify Reddit posts that contain misinformation, we then applied the BTM to these posts. We performed stemming for each word to remove morphological affixes (eg, *dies* to *die* and *denied* to *deni*). The number of topics is a hyperparameter for the BTM, which determines the number of topics that will be generated. We evaluated 5, 10, and 20 as the number of topics and selected 10 through a manual review of the topics and associated posts. We then manually reviewed these topics, associated words, and posts to further identify relevant topics associated with vaccine misinformation. The BTM also outputs the prevalence of each identified topic. Word clouds were then adopted to provide a graphic representation of these topics, where the size of each word is proportional to its probability of appearing in posts about that topic [54]. To examine the association of the identified topics, we further performed network analysis among these topics.

Ethics Approval and Consent to Participate

This study received institutional review board exemption from the committee for the protection of human subjects at the University of Texas Health Science Center at Houston. The reference number is HSC-SBMI-20-0151.

Results

Misinformation Annotation

In total, 28,121 Reddit posts were collected from 2007 to 2017 from more than 16,633 unique users. The statistics of these posts as well as their distributions in subreddits (ie, user-created discussion boards where posts are organized by a subject) are shown in Table 1. There was an increasing trend of HPV vaccine–related discussions (in terms of both the number of posts and number of unique users) during the study period. There were 207,651 upvotes (a user likes the post) and 10,700 downvotes (a user does not like the post) for these posts. Of the 28,121 posts, we manually labeled 2200 (7.82%) randomly selected posts. We measured the annotation agreement by calculating the Cohen κ among the 3 annotators: 0.5578 for JD and HS, 0.5216 for JD and SP, and 0.4685 for HS and SP. The agreement scores are considered moderate according to El Eman [32], which indicates a good quality of our gold standard. Among these 2200 posts, 396 (18%) were annotated as vaccine misinformation, whereas 1804 (82%) were annotated as nonmisinformation. The highly imbalanced label distribution created barriers to achieving high performance for the classification algorithms.

Table 1. The statistics of the human papillomavirus Reddit posts corpus. For statistics regarding Reddit users, we removed the posts if the accounts were unavailable.

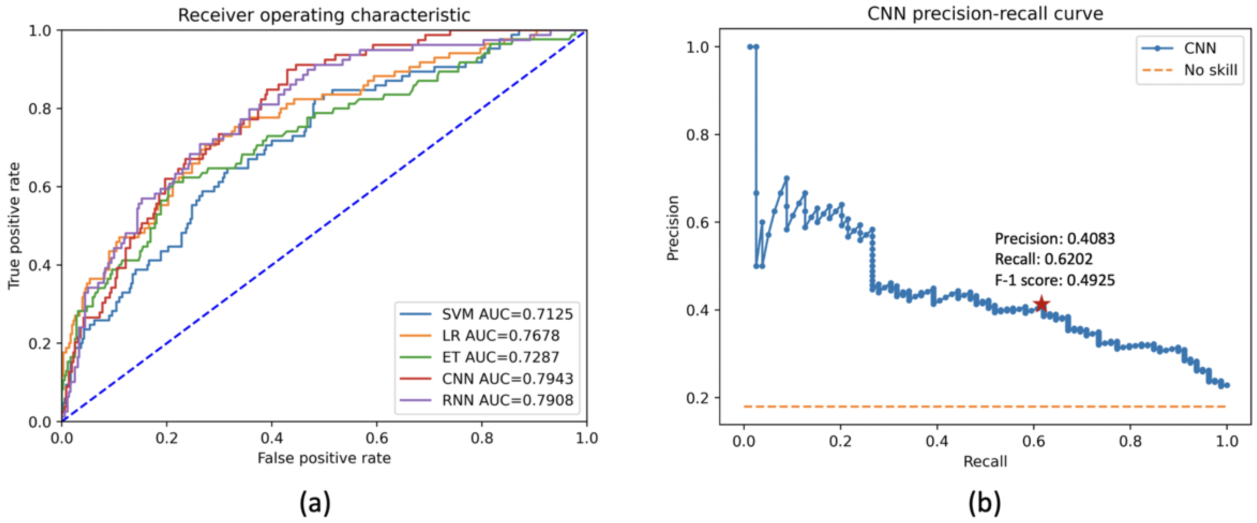
Year	Total posts	Total up-votes	Total down-votes	Total unique users	User posts distribution, mean (SD)	Most frequent subreddits[a] (top 3)	Subreddit post distribution, mean (SD)
2007	15	51	1	10	1.00 (0.00)	Reddit, 11; politics, 3; science, 1	5.00 (5.29)
2008	172	335	35	100	1.34 (1.32)	Reddit, 57; science, 54; health, 23	11.47 (18.88)
2009	414	1563	206	249	1.39 (0.93)	Science, 81; AskReddit, 51; Reddit, 47	14.28 (18.96)
2010	546	1655	155	346	1.33 (0.79)	AskReddit, 95; sex, 83; TwoXChromosomes, 72	12.13 (22.37)
2011	2156	12,711	927	1382	1.37 (1.47)	Politics, 298; TwoXChromosomes, 207; AskReddit, 203	19.42 (48.48)
2012	2457	12,812	739	1641	1.32 (1.34)	AskReddit, 457; TwoXChromosomes, 308; sex, 221	13.96 (48.60)
2013	3864	26,623	1416	2540	1.39 (2.44)	Science, 490; AskReddit, 375; sex, 297	15.97 (51.53)
2014	3488	21,562	1581	2348	1.39 (2.62)	Sex, 325; AskReddit, 292; science, 291	10.67 (34.81)
2015	4714	35,801	1761	3383	1.38 (1.61)	News, 378; science, 357; AskReddit, 347	11.67 (41.28)
2016	4417	38,123	1436	3137	1.39 (1.43)	AskReddit, 378; TwoXChromosomes, 262; sex, 255	10.44 (34.25)
2017	5878	56,415	2443	3752	1.56 (9.09)	AskReddit, 446; sex, 402; news, 387	11.28 (40.95)

[a] Numbers included indicate counts.

Misinformation Detection and Classification

The LR algorithm demonstrated the highest AUC value (0.7678) among the 3 traditional ML algorithms used to identify vaccine misinformation in the Reddit posts (Figure 2). Both DL algorithms (CNN and RNN) achieved higher AUC values than the traditional ML algorithms. The CNN model slightly outperformed the RNN model (0.7943 vs 0.7908) in the identification of misinformation. The CNN model with the optimal cutoff was applied to classify the Reddit posts that contained vaccine misinformation. The precision and recall curves of the CNN model are shown in Figure 2. The optimal cutoff led to a precision of 0.4083, a recall of 0.6202, and an F_1 score of 0.4925. Together with 1.41% (396/28,121) of the Reddit posts that were manually annotated as misinformation, 25.63% (7207/28,121) of the random subset of posts were classified as vaccine misinformation.

Figure 2. The performance of machine learning algorithms on human papillomavirus misinformation identification. (a) Receiver operating characteristic and (b) convolutional neural networks precision-recall curves. AUC: area under the curve; ET: extremely randomized trees; CNN: convolutional neural networks; LR: logistic regression; RNN: recurrent neural network; SVM: support vector machines.



Topic modeling generated 10 topics from 7207 Reddit posts that were classified as vaccine misinformation. Through qualitative analysis of these 10 algorithm-identified topics and a review of their relevant Reddit posts, we condensed them into 7 (6 major topics + *Other*) topics. The word clouds for the 6 identified misinformation topics are shown in Figure S1 of Multimedia Appendix 1. The 6 major topics, the percentage of posts assigned to the topic, and post examples are listed in Table 2.

Table 2. The major topics of misinformation identified by topic modeling (n=7207).

Misinformation topic	Prevalence, n (%)	Explanation	Examples (excerpts)
General vaccine adverse events	2672 (37.07)	Promotion of general misinformation regarding the safety of HPV[a] vaccine	• "The HPV vaccine is unnecessary and unsafe." • "The HPV vaccination causes retardation."
Conspiracy theory	1072 (14.87)	Propagation of conspiracy theories about HPV vaccine and fraud by the government and large pharmaceutical companies (eg, Merck)	• "Rick Perry signed an executive order trying to mandate Gardasil vaccines for young girls. He did this to make money for his buddies at Merck." • "The HPV vaccine is a joke pressed on schools and administrations by greedy pharmaceutical corporations."
Citing unfounded studies	989 (13.72)	This type of misinformation can be very misleading because it tends to cite and interpret *scientific* studies from sources that are not scientifically peer reviewed	• "Flawed safety study on HPV vaccine triggers butterfly effect - leaked e-mails reveal who suppressed info on dangerous particles in vaccine" • "We find that HPV vaccine clinical trials design, and data interpretation of both efficacy and safety outcomes, were largely inadequate"
Vaccine deaths and serious reactions	520 (7.21)	Propagation of HPV vaccine–induced death and serious adverse reactions	• "For instance, the HPV vaccine has caused children to die." • "The shots have killed women. The adverse event reports from the FDA[b] about HPV vaccination read like a litany of horrors."
Aluminum-containing adjuvants	456 (6.33)	Promoting misinformation on safety issues of aluminum-containing adjuvants in vaccines	• "Another good keyword is for Gardasil adverse drug reactions. That's how I found this: 'each 0.5-ml dose of the vaccine contains approximately 225 mcg of aluminum (as amorphous aluminum hydroxyphosphate sulfate adjuvant)' This study clearly shows that aluminum found in vaccines can cause neurologic damage." • "Brain damage and autoimmune diseases can be caused by aluminum adjuvants. Aluminum adjuvant is in the HPV vaccine."
Vaccine and autism	198 (2.75)	Promoting misinformation on the discredited links between vaccine and autism	• "The HPV vaccine causes autism" • "Vaccine court awarded millions to two children with autism. CDC[c] report creates controversy for Merck's Gardasil vaccine"

[a] HPV: human papillomavirus.

[b] FDA: Food and Drug Administration.

[c] CDC: US Centers for Disease Control and Prevention.

Misinformation Network Analysis

We further analyzed the network among the identified topics. For each Reddit post, we identified the 2 most associated topics (ranked by probability generated by the BTM). We assume that these top 2 topics were linked for that post, which was considered an undirected edge in the network. Figure S2 in Multimedia Appendix 1 shows the misinformation topic network among these 7 topics. The size of the circle for each topic is proportional to the degree of associations of the topic (ie, the number of connections with other topics). The width of the edge is proportional to the number of connections between the 2 topics.

Discussion

Principal Findings

In this study, we evaluated the use of different ML-based approaches to analyze Reddit discussions related to HPV vaccine misinformation. The CNN and RNN algorithms improved the AUC value compared with the traditional ML algorithms. A BTM was adopted to further explore the major topics related to vaccine misinformation discussions. Overall, 6 major topics related to HPV vaccine misinformation, including *Vaccine death and serious reactions* and *Aluminum-containing adjuvants* were identified. The *Vaccine adverse effect*, which refers to general misinformation regarding safety issues, is the most prevalent topic within HPV vaccine misinformation.

The highest proportion of vaccine misinformation content on Reddit identified with our approach concerned general vaccine adverse effects (2672/7207, 37.07%), followed by content about vaccine conspiracy theories (1072/7207, 14.87%). These results are consistent with previous analyses of social media–based vaccine misinformation, which found that inaccuracies about vaccine knowledge and risk (37.9%) made up most of the social media posts with negative vaccine sentiments [55]. The same small study also found that 13.8% of the negative posts about vaccines included distrust of government and pharmaceutical companies, which closely mirrors our findings in a larger sample from Reddit.

We further analyzed the top subreddits among the posts that ML inferred as containing misinformation and nonmisinformation. The subreddits that contain the most misinformation-related posts include *science* (n=653), *AskReddit* (n=604), *conspiracy* (n=593), and *politics* (n=397). On the contrary, the subreddits that contain the most nonmisinformation-related posts include *AskReddit* (n=2040), *sex* (n=1994), *TwoXChromosomes* (n=1652), and *science* (n=1385). Besides general popular subreddits such as *science* and *AskReddit*, misinformation tends to cluster in subreddits such as *conspiracy* and *politics*. There is an increasing trend of discussing HPV vaccine–related topics on Reddit from 15 posts in 2007 to 5878 posts in 2017. There is a decreasing trend in the proportion of misinformation over time on Reddit. The proportion of misinformation ranged from 41.8% (72/172; in year 2008) to 53% (8/15; in year 2007) during the period 2007 to 2009, whereas it ranged from 22.84% (1009/4417; in year 2016) to 33.85% (730/2156; in year 2011) during the period 2010 to 2017. The decrease could be a result of the continuous promotion efforts made by public health professionals, as well as an increase in internet verification skills among users.

The results of our network analysis of the 6 identified vaccine misinformation topics (in addition to *Other*) further reinforce our findings and demonstrate the strength of the connectedness of each topic. Although general concerns about the safety of the vaccine emerged as the main source of hesitancy regarding HPV vaccination, the network analysis indicates that the other prominent topics identified, such as the presence of conspiracy theories, may also be rooted in fears about the side effects of the vaccine. Mere exposure to beliefs that the government and pharmaceutical companies gain or profit from mass vaccination through deception or at the consumers' expense, has strong negative effects on attitudes about the safety and effectiveness of vaccines, consequently affecting choices about whether to vaccinate [56].

Of note, the annotators anecdotally observed that the Reddit posts identified in this study did not seem to be connected to any organized movements; rather, they were by single users advocating their personal views. A potential method to combat these misinformed messages once identified is to counter them with an organized campaign, composed of factual, evidence-based messages, that does not acknowledge disinformation. As other studies have noted, acknowledging and deferring to web-based disputes related to vaccines may cause health information seekers to doubt established evidence regarding vaccine efficacy and safety [57]. In addition, it has been found that attempting to correct misinformation directly often reinforces the sentiments of those holding strong antivaccine views [58].

To the best of our knowledge, this is one of the early efforts to explore the use of *automated* ML algorithms (eg, ML and NLP) to identify and classify HPV vaccine misinformation in social media discussions. We chose the HPV vaccine as the use case for our analysis, but the proposed methodology framework can also be applied to other types of vaccines or other pertinent health-related topics. The ML-based framework is also scalable to big social media data. Our work could assist policy makers and the industry to accurately understand and address the spread of health misinformation on social media. The methodology framework developed in this study is generalizable to other social media platforms such as Twitter and can be used to identify misinformation in both retrospective and real-time social media feeds. The use of this methodology could be incorporated into social media platforms dedicated to curbing the spread of health-related misinformation on these sites, although the ethical ramifications of such restrictions should be taken into consideration.

Limitations and Future Work

This study should be interpreted in light of limitations and future research needs. Given the unique features of social media language, the accurate identification of misinformation is a very challenging task. The best algorithm achieved an AUC value of 0.7943, and there is some room for improving this performance. Our current ML classifier has a higher recall than precision (0.6202 vs 0.4083). This means that the classifier tends to label both misinformation and nonmisinformation as misinformation. In a real-world scenario, the classifier may serve as a tool to prescreen misinformation, and more rigorous fact-checking methods (eg, human checking) would be needed to label true misinformation posts. The high imbalanced label distribution (ie, only 396/2200, 18%) of the posts were labeled as misinformation in the gold standard corpus) hurt the ML algorithm because most of the ML algorithms used for classification were designed based on the assumption of an equal number of examples for each class. Imbalanced label distribution results in models that have poor predictive performance, specifically for the minority class (eg, misinformation in our case) [59]. As we further refine and expand the gold standard corpus, which is critical for the evaluation and training of ML algorithms, we expect the performance to improve. In addition, we will explore the use of data augmentation techniques [60] and random oversampling methods [61,62] to alleviate the issues caused by imbalanced label distribution. Other emerging advanced DL algorithms such as Bidirectional Encoder Representations from Transformers (BERT) [63] hold promise for improved performance. In addition, we performed annotation at the level of Reddit posts, which may have sacrificed precision. A single Reddit post often contains multiple sentences, allowing a mix of misinformation and nonmisinformation to exist in a single post. Therefore, a Reddit post annotated as *misinformation* could also contain evidence-based facts. Future research can establish the effect of annotation and classification at the sentence level to improve the precision of misinformation identification. In addition, the identification of *abstract topics* from topic modeling is a semiautomated process combined with expert review. However, topic assignment and summarization may be subjective and suffer from biases as well. In future, we can explore the use of supervised algorithms for more precise topic discovery.

Conclusions

Our ML-based approaches demonstrated efficacy in the automated identification and classification of HPV vaccine misinformation in discussions on the social media platform Reddit. The large quantity of web- and social media–based medical and public health information available may make it difficult for those with low health and web literacy to navigate and find authentic and evidence-based information. Although our ML algorithm does not solve the problem of health and vaccine misinformation single-handedly, we provide an innovative stepping stone that may bridge multiple approaches for combating this invasive and growing public health concern. The accurate and timely understanding of vaccine misinformation on social media can assist vaccine promotion campaigns to prevent such information from misleading the vulnerable public. Our methodology could also be applied to other social media platforms such as Twitter, although new labeled data would be necessary.

Acknowledgments

This research was supported by the National Institutes of Health under award numbers 2R01LM010681-05, R01 LM011829, and NCI T32/CA057712 (SP); and the Cancer Prevention and Research Institute of Texas training grant #RP160015. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Library of Medicine and the Cancer Prevention and Research Institute of Texas.

Authors' Contributions

JD and CT have full access to all data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Study concept and design was conducted by JD, SP, MA, and CT. Data annotation was carried out by JD, SP, and HS. JD, SP, RS, MA, and CT were involved in drafting of the manuscript. Acquisition, analysis, or interpretation of data was conducted by JD, SP, MA, and CT. Critical revision of the manuscript for important intellectual content was carried out by all authors. Study supervision was conducted by CT.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplemental tables and figures.
[DOCX File , 1055 KB - jmir_v23i8e26478_app1.docx]

References

1. Human Papillomavirus (HPV) - reasons to get vaccinated. Centers for Disease Control and Prevention. 2019. URL: https://www.cdc.gov/hpv/parents/vaccine/six-reasons.html [accessed 2021-07-02]
2. Saraiya M, Unger E, Thompson T, Lynch CF, Hernandez BY, Lyu CW, HPV Typing of Cancers Workgroup. US assessment of HPV types in cancers: implications for current and 9-valent HPV vaccines. J Natl Cancer Inst 2015 Jun;107(6):djv086 [FREE Full text] [doi: 10.1093/jnci/djv086] [Medline: 25925419]
3. HPV vaccine: who needs it, how it works. Mayo Clinic. 2019. URL: https://www.mayoclinic.org/diseases-conditions/hpv-infection/in-depth/hpv-vaccine/art-20047292 [accessed 2020-02-09]
4. Zimet GD, Rosberger Z, Fisher WA, Perez S, Stupiansky NW. Beliefs, behaviors and HPV vaccine: correcting the myths and the misinformation. Prev Med 2013 Nov;57(5):414-418 [FREE Full text] [doi: 10.1016/j.ypmed.2013.05.013] [Medline: 23732252]
5. Etter DJ, Zimet GD, Rickert VI. Human papillomavirus vaccine in adolescent women: a 2012 update. Curr Opin Obstet Gynecol 2012 Oct;24(5):305-310. [doi: 10.1097/GCO.0b013e3283567005] [Medline: 22781077]
6. Holman DM, Benard V, Roland KB, Watson M, Liddon N, Stokley S. Barriers to human papillomavirus vaccination among US adolescents: a systematic review of the literature. JAMA Pediatr 2014 Jan 01;168(1):76-82 [FREE Full text] [doi: 10.1001/jamapediatrics.2013.2752] [Medline: 24276343]
7. Poland G, Jacobson R. Understanding those who do not understand: a brief review of the anti-vaccine movement. Vaccine 2001 Mar;19(17-19):2440-2445. [doi: 10.1016/s0264-410x(00)00469-2]
8. Geoghegan S, O'Callaghan KP, Offit PA. Vaccine safety: myths and misinformation. Front Microbiol 2020 Mar 17;11:372 [FREE Full text] [doi: 10.3389/fmicb.2020.00372] [Medline: 32256465]
9. 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]
10. Kata A. A postmodern Pandora's box: anti-vaccination misinformation on the internet. Vaccine 2010 Feb 17;28(7):1709-1716. [doi: 10.1016/j.vaccine.2009.12.022] [Medline: 20045099]

11. Anderson J, Rainie L. Experts optimistic about the next 50 years of digital life. Pew Research Center. 2019. URL: https://www.pewresearch.org/internet/2019/10/28/experts-optimistic-about-the-next-50-years-of-digital-life/ [accessed 2021-07-16]
12. Levin S. Facebook promised to tackle fake news. But the evidence shows it's not working. The Guardian. 2017. URL: https://www.theguardian.com/technology/2017/may/16/facebook-fake-news-tools-not-working [accessed 2020-02-10]
13. Ghosh D, Scott B. Disinformation is becoming unstoppable. Time. 2018. URL: https://time.com/5112847/facebook-fake-news-unstoppable/ [accessed 2020-02-10]
14. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
15. Chou WS, Oh A, Klein WM. Addressing health-related misinformation on social media. J Am Med Assoc 2018 Dec 18;320(23):2417-2418. [doi: 10.1001/jama.2018.16865] [Medline: 30428002]
16. Qazvinian V, Rosengren E, Radev D, Mei Q. Rumor has it: identifying misinformation in microblogs. In: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing. 2011 Presented at: Conference on Empirical Methods in Natural Language; July, 2011; Edinburgh, Scotland, UK p. 1589-1599 URL: https://aclanthology.org/D11-1147
17. Hamidian S, Diab M. Rumor detection and classification for Twitter data. arXiv. URL: http://arxiv.org/abs/1912.08926 [accessed 2021-07-02]
18. Machine learning. Wikipedia. URL: https://en.wikipedia.org/wiki/Machine_learning [accessed 2020-02-09]
19. Ghenai A, Mejova Y. Catching Zika fever: application of crowdsourcing and machine learning for tracking health misinformation on Twitter. In: Proceedings of the IEEE International Conference on Healthcare Informatics (ICHI). 2017 Presented at: IEEE International Conference on Healthcare Informatics (ICHI); Aug. 23-26, 2017; Park City, UT, USA. [doi: 10.1109/ichi.2017.58]
20. Kinsora A, Barron K, Mei Q, Vydiswaran V. Creating a labeled dataset for medical misinformation in health forums. In: Proeedings of the IEEE International Conference on Healthcare Informatics (ICHI). 2017 Presented at: IEEE International Conference on Healthcare Informatics (ICHI); Aug. 23-26, 2017; Park City, UT, USA p. 456-461. [doi: 10.1109/ichi.2017.93]
21. Du J, Xu J, Song H, Liu X, Tao C. Optimization on machine learning based approaches for sentiment analysis on HPV vaccines related tweets. J Biomed Semantics 2017 Mar 03;8(1):9 [FREE Full text] [doi: 10.1186/s13326-017-0120-6] [Medline: 28253919]
22. Zhou X, Coiera E, Tsafnat G, Arachi D, Ong M, Dunn A. Using social connection information to improve opinion mining: identifying negative sentiment about HPV vaccines on Twitter. Stud Health Technol Inform 2015;216:761-765. [doi: 10.3233/978-1-61499-564-7-761]
23. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015 May 28;521(7553):436-444. [doi: 10.1038/nature14539] [Medline: 26017442]
24. Mediakix. 2018. URL: https://mediakix.com/blog/reddit-statistics-users-demographics/ [accessed 2020-02-05]
25. Frequently asked questions. National Institute for Communicable Diseases. URL: https://www.nicd.ac.za/diseases-a-z-index/covid-19/frequently-asked-questions/ [accessed 2018-12-04]
26. Barnard M, George P, Perryman ML, Wolff LA. Human papillomavirus (HPV) vaccine knowledge, attitudes, and uptake in college students: implications from the Precaution Adoption Process Model. PLoS One 2017 Aug 7;12(8):e0182266 [FREE Full text] [doi: 10.1371/journal.pone.0182266] [Medline: 28786994]
27. Thompson EL, Vamos CA, Vázquez-Otero C, Logan R, Griner S, Daley EM. Trends and predictors of HPV vaccination among U.S. college women and men. Prev Med 2016 May;86:92-98. [doi: 10.1016/j.ypmed.2016.02.003] [Medline: 26868093]
28. Abrol E, Groszmann M, Pitman A, Hough R, Taylor RM, Aref-Adib G. Exploring the digital technology preferences of teenagers and young adults (TYA) with cancer and survivors: a cross-sectional service evaluation questionnaire. J Cancer Surviv 2017 Dec;11(6):670-682 [FREE Full text] [doi: 10.1007/s11764-017-0618-z] [Medline: 28634734]
29. Sedgwick P. Receiver operating characteristic curves. Br Med J 2011 Jul 13;343(jul13 2):d4302. [doi: 10.1136/bmj.d4302]
30. Pushshift.io. URL: https://pushshift.io/ [accessed 2021-07-16]
31. Amith M, Tao C. Representing vaccine misinformation using ontologies. J Biomed Semantics 2018 Aug 31;9(1):22 [FREE Full text] [doi: 10.1186/s13326-018-0190-0] [Medline: 30170633]
32. Emam EK. Benchmarking Kappa: interrater agreement in software process assessments. Empir Softw Eng 1999;4(2):113-133. [doi: 10.1023/A:1009820201126]
33. Jurafsky D, Martin JH. Speech and language processing 2nd Edition. Upper Saddle River, New Jersey: Prentice-Hall, Inc; 2009.
34. Naz S, Sharan A, Malik N. Sentiment classification on Twitter data using support vector machine. In: Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence (WI). 2018 Presented at: IEEE/WIC/ACM International Conference on Web Intelligence (WI); Dec. 3-6, 2018; Santiago, Chile. [doi: 10.1109/wi.2018.00-13]
35. Zhang W, Yoshida T, Tang X. Text classification based on multi-word with support vector machine. Knowledge-Based Systems 2008 Dec;21(8):879-886. [doi: 10.1016/j.knosys.2008.03.044]
36. Tong S, Koller D. Support vector machine active learning with applications to text classification. J Mach Learn Res. (Nov) 2001;2:45-66 [FREE Full text]

37. Pranckevičius T, Marcinkevičius V. Comparison of naive bayes, random forest, decision tree, support vector machines, and logistic regression classifiers for text reviews classification. Balt J Mod Comput 2017;5(2):221. [doi: 10.22364/bjmc.2017.5.2.05]
38. Aborisade O, Anwar M. Classification for authorship of tweets by comparing logistic regression and naive bayes classifiers. In: Proceedings of the IEEE International Conference on Information Reuse and Integration (IRI). 2018 Presented at: IEEE International Conference on Information Reuse and Integration (IRI); July 6-9, 2018; Salt Lake City, UT, USA. [doi: 10.1109/iri.2018.00049]
39. Du J, Luo C, Shegog R, Bian J, Cunningham RM, Boom JA, et al. Use of deep learning to analyze social media discussions about the human papillomavirus vaccine. JAMA Netw Open 2020 Nov 02;3(11):e2022025 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.22025] [Medline: 33185676]
40. Du J, Cunningham RM, Xiang Y, Li F, Jia Y, Boom JA, et al. Leveraging deep learning to understand health beliefs about the Human Papillomavirus Vaccine from social media. NPJ Digit Med 2019 Apr 15;2(1):27. [doi: 10.1038/s41746-019-0102-4]
41. Ramos J. Using TF-IDF to determine word relevance in document queries. In: Proceedings of the First Instructional Conference on Machine Learning. 2003 Presented at: First Instructional Conference on Machine Learning; June 23-24, 2003; Los Angeles, California p. 29-48 URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.121.1424&rep=rep1&type=pdf
42. Du J, Luo C, Wei Q, Chen Y, Tao C. Exploring difference in public perceptions on HPV vaccine between gender groups from Twitter using deep learning. arXiv. 2019. URL: http://arxiv.org/abs/1907.03167 [accessed 2021-07-02]
43. Zhou P, Shi W, Tian J, Qi Z, Li B, Hao H, et al. Attention-based bidirectional long short-term memory networks for relation classification. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. 2016 Presented at: 54th Annual Meeting of the Association for Computational Linguistics; August, 2016; Berlin, Germany p. 207-212. [doi: 10.18653/v1/p16-2034]
44. Kharde VA, Sonawane SS. Sentiment Analysis of Twitter Data: A Survey of Techniques. IJCA 2016 Apr 15;139(11):5-15. [doi: 10.5120/ijca2016908625]
45. Young T, Hazarika D, Poria S, Cambria E. Recent trends in deep learning based natural language processing [Review article]. IEEE Comput Intell Mag 2018 Aug;13(3):55-75. [doi: 10.1109/MCI.2018.2840738]
46. Zhang L, Wang S, Liu B. Deep learning for sentiment analysis: a survey. WIREs Data Mining Knowl Discov 2018 Mar 30;8(4):e1253. [doi: 10.1002/widm.1253]
47. Zhang X, Zhao J, Lecun Y. Character-level convolutional networks for text classification. In: Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press; 2015:649-657 URL: https://proceedings.neurips.cc/paper/2015/file/250cf8b51c773f3f8dc8b4be867a9a02-Paper.pdf
48. Liu G, Guo J. Bidirectional LSTM with attention mechanism and convolutional layer for text classification. Neurocomputing 2019 Apr;337:325-338. [doi: 10.1016/j.neucom.2019.01.078]
49. Liu X, You X, Zhang X, Wu J, Lv P. Tensor graph convolutional networks for text classification. arXiv. 2020. URL: https://arxiv.org/abs/2001.05313 [accessed 2021-07-02]
50. Pennington J, Socher R, Manning C. GloVe: Global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014 Presented at: EMNLP Conference on Empirical Methods in Natural Language Processing; October, 2014; Doha, Qatar p. 1532-1543. [doi: 10.3115/v1/d14-1162]
51. Tokenizer. spaCy API Documentation. URL: https://spacy.io/api/tokenizer [accessed 2020-05-29]
52. Yan X, Guo J, Lan Y, Cheng X. A biterm topic model for short texts. In: Proceedings of the 22nd International Conference on World Wide Web. 2013 Presented at: WWW '13: 22nd International World Wide Web Conference; May 13-17, 2013; Rio de Janeiro, Brazil p. 1445-1456. [doi: 10.1145/2488388.2488514]
53. xiaohuiyan/BTM: Code for Biterm Topic Model. Github. URL: https://github.com/xiaohuiyan/BTM [accessed 2020-03-05]
54. Generating word cloud in Python. GeeksforGeeks. URL: https://www.geeksforgeeks.org/generating-word-cloud-python/ [accessed 2020-06-08]
55. Tustin JL, Crowcroft NS, Gesink D, Johnson I, Keelan J, Lachapelle B. User-driven comments on a Facebook advertisement recruiting Canadian parents in a study on immunization: content analysis. JMIR Public Health Surveill 2018 Sep 20;4(3):e10090 [FREE Full text] [doi: 10.2196/10090] [Medline: 30249585]
56. Jolley D, Douglas KM. The effects of anti-vaccine conspiracy theories on vaccination intentions. PLoS One 2014 Feb 20;9(2):e89177 [FREE Full text] [doi: 10.1371/journal.pone.0089177] [Medline: 24586574]
57. Dixon G, Clarke C. The effect of falsely balanced reporting of the autism-vaccine controversy on vaccine safety perceptions and behavioral intentions. Health Educ Res 2013 Apr 27;28(2):352-359. [doi: 10.1093/her/cys110] [Medline: 23193194]
58. Nyhan B, Reifler J. When corrections fail: the persistence of political misperceptions. Polit Behav 2010 Mar 30;32(2):303-330. [doi: 10.1007/s11109-010-9112-2]
59. Krawczyk B. Learning from imbalanced data: open challenges and future directions. Prog Artif Intell 2016 Apr 22;5(4):221-232. [doi: 10.1007/s13748-016-0094-0]
60. Wei J, Zou K. EDAasy data augmentation techniques for boosting performance on text classification tasks. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference

on Natural Language Processing (EMNLP-IJCNLP). 2019 Presented at: Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP); November, 2019; Hong Kong, China p. 6382-6388. [doi: 10.18653/v1/d19-1670]
61. Moreo A, Esuli A, Sebastiani F. Distributional random oversampling for imbalanced text classification. In: Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2016 Presented at: SIGIR '16: The 39th International ACM SIGIR Conference on Research and Development in Information Retrieval; July 17-21, 2016; Pisa Italy p. 805-808. [doi: 10.1145/2911451.2914722]
62. Charte F, Rivera AJ, del Jesus MJ, Herrera F. Addressing imbalance in multilabel classification: measures and random resampling algorithms. Neurocomputing 2015 Sep;163:3-16. [doi: 10.1016/j.neucom.2014.08.091]
63. Devlin J, Chang M, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805. 2018 Oct 11. URL: https://arxiv.org/abs/1810.04805

Abbreviations

AUC: area under the receiver operating characteristic curve
BERT: Bidirectional Encoder Representations from Transformers
BTM: Biterm Topic model
CNN: convolutional neural network
DL: deep learning
GloVe: Global Vectors for Word Representation
HPV: human papillomavirus
LR: logistic regression
ML: machine learning
NLP: natural language processing
RNN: recurrent neural network
TF-IDF: term frequency-inverse document frequency
VAXMO: Vaccine Misinformation Ontology

Edited by R Kukafka; submitted 13.12.20; peer-reviewed by X Jing, Z Zhang, S Kim; comments to author 16.02.21; revised version received 14.04.21; accepted 06.05.21; published 05.08.21.

Please cite as:
Du J, Preston S, Sun H, Shegog R, Cunningham R, Boom J, Savas L, Amith M, Tao C
Using Machine Learning–Based Approaches for the Detection and Classification of Human Papillomavirus Vaccine Misinformation: Infodemiology Study of Reddit Discussions
J Med Internet Res 2021;23(8):e26478
URL: https://www.jmir.org/2021/8/e26478
doi:10.2196/26478
PMID:34383667

©Jingcheng Du, Sharice Preston, Hanxiao Sun, Ross Shegog, Rachel Cunningham, Julie Boom, Lara Savas, Muhammad Amith, Cui Tao. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Mining of Opinions on COVID-19 Large-Scale Social Restrictions in Indonesia: Public Sentiment and Emotion Analysis on Online Media

Andi Muhammad Tri Sakti[1,2], MSSc; Emma Mohamad[1,2], PhD; Arina Anis Azlan[1,2], PhD

[1]Centre for Research in Media and Communication, Faculty of Social Sciences and Humanities, Universiti Kebangsaan Malaysia, Bangi, Malaysia

[2]UKM x UNICEF Communication for Development Centre in Health, Faculty of Social Sciences and Humanities, Universiti Kebangsaan Malaysia, Bangi, Malaysia

Corresponding Author:
Arina Anis Azlan, PhD
Centre for Research in Media and Communication
Faculty of Social Sciences and Humanities
Universiti Kebangsaan Malaysia
UKM Bangi
Bangi, 43600
Malaysia
Phone: 60 0123065261
Email: arina@ukm.edu.my

Abstract

Background: One of the successful measures to curb COVID-19 spread in large populations is the implementation of a movement restriction order. Globally, it was observed that countries implementing strict movement control were more successful in controlling the spread of the virus as compared with those with less stringent measures. Society's adherence to the movement control order has helped expedite the process to flatten the pandemic curve as seen in countries such as China and Malaysia. At the same time, there are countries facing challenges with society's nonconformity toward movement restriction orders due to various claims such as human rights violations as well as sociocultural and economic issues. In Indonesia, society's adherence to its large-scale social restrictions (LSSRs) order is also a challenge to achieve. Indonesia is regarded as among the worst in Southeast Asian countries in terms of managing the spread of COVID-19. It is proven by the increased number of daily confirmed cases and the total number of deaths, which was more than 6.21% (1351/21,745) of total active cases as of May 2020.

Objective: The aim of this study was to explore public sentiments and emotions toward the LSSR and identify issues, fear, and reluctance to observe this restriction among the Indonesian public.

Methods: This study adopts a sentiment analysis method with a supervised machine learning approach on COVID-19-related posts on selected media platforms (Twitter, Facebook, Instagram, and YouTube). The analysis was also performed on COVID-19-related news contained in more than 500 online news platforms recognized by the Indonesian Press Council. Social media posts and news originating from Indonesian online media between March 31 and May 31, 2020, were analyzed. Emotion analysis on Twitter platform was also performed to identify collective public emotions toward the LSSR.

Results: The study found that positive sentiment surpasses other sentiment categories by 51.84% (n=1,002,947) of the total data (N=1,934,596) collected via the search engine. Negative sentiment was recorded at 35.51% (686,892/1,934,596) and neutral sentiment at 12.65% (244,757/1,934,596). The analysis of Twitter posts also showed that the majority of public have the emotion of "trust" toward the LSSR.

Conclusions: Public sentiment toward the LSSR appeared to be positive despite doubts on government consistency in executing the LSSR. The emotion analysis also concluded that the majority of people believe in LSSR as the best method to break the chain of COVID-19 transmission. Overall, Indonesians showed trust and expressed hope toward the government's ability to manage this current global health crisis and win against COVID-19.

(J Med Internet Res 2021;23(8):e28249) doi:10.2196/28249

KEYWORDS

large-scale social restrictions; social media; public sentiment; Twitter; COVID-19; infodemiology; infoveillance

Introduction

Social quarantine measures, such as movement control order, shelter in place, and lockdowns, are considered effective methods to curb the spread of COVID-19 [1]. Countries all over the globe affected by this pandemic have at least considered implementing this modus as a national strategy to bring COVID-19 cases down and reduce the burden on their health care system. Until a vaccine can be effectively administered to the public, the only effort that can be done to reduce virus transmission sustainably is by exercising prevention measures and embracing it as the new living norm. This includes restricting movement that reduces social contact and improves physical distancing [2]. Physical distancing is performed to decrease the risk of virus transmission from one who is likely to be infected to other healthy individuals [3]. A study measuring the effectiveness of physical distancing in England found that this method can significantly decrease the contact-level frequency up to 74% and was the key factor in decreasing the total confirmed cases in the country [4]. Effective physical distancing correlates with a gradual decline in the number of cases infected [5].

In most countries affected by COVID-19, national-level physical distancing order such as lockdowns, movement control, and shelter in place are implemented as strategies to bring down the number of coronavirus cases. This regulation has been proven as the most effective strategy to eradicate the spread of the virus in several countries such as Brunei [6], New Zealand, and Vietnam [7]. As a result, those 3 countries have succeeded in flattening the curve since the first wave of infection up until now. A valuable lesson can also be learned from Indonesia's closest neighboring country—Malaysia. Through a study on COVID-19 control measures, Ng et al [8] summarized that the movement control order had managed to suppress the transmission within only 3 weeks after its implementation. Thus, despite difficulties in implementing physical distancing measures in the long run, this method is a critical and effective measure in curbing the spread of infection [9].

After the first case was reported on March 2, 2020, the COVID-19 curve of Indonesia started to show a significant increase at the end of the month with no signs of decline, as the number of infected cases continued to hit a new peak every day. In that month, with 136 deaths out of 1528 total positive cases, Indonesia surpassed China in its case fatality rate (8.90%). China's death rate was at 4% at the time [10]. The implementation of the large-scale social restriction (LSSR) on March 31, 2020, was overdue, as the number of cases was significantly high [11]. At the time, the number of positive cases in Indonesia was still 2 times lower compared with Malaysia, but as of early June 2020, the country had the second highest number of positive confirmed cases in Southeast Asia after Singapore. Indonesia became the country with the highest fatality rate in Asia, with a case fatality rate of 5.94% (n=1573) among the overall confirmed cases (N=26,473) [12]. If social restriction is a reference to the success of a country in handling and suppressing the COVID-19 outbreak, then the high number of daily confirmed cases reported in Indonesia indicates the poor implementation of the policy and government supervision of the LSSR. According to The Institute for Development of Economics and Finance [13], Indonesia is the least successful country in implementing the movement restriction policy among Southeast Asian countries. The number of confirmed cases in Indonesia is considered the worst compared with neighboring countries.

LSSR implementation in Indonesia indeed has faced many obstacles since its first inception. The success of LSSR is not merely determined by policy execution by the government, but is significantly influenced by public willingness to participate by voluntarily staying at home and not visiting public places. Therefore, it is useful to examine public response toward the implementation of LSSR. However, studies on public response toward LSSR effectiveness are limited. Most prior public-oriented studies have focused on COVID-19 knowledge, attitudes, and behaviors [14-19]. Furthermore, only few studies discussing public response toward social quarantine policy, especially in Indonesia, are available.

Accordingly, this study aims to identify public response toward government policy, specifically the LSSR, in Indonesia. Studies on public response toward policy implementation regarding COVID-19 prevention are imperative, particularly when the government as the policy maker is expected to be sensitive to societal needs and behavior [20]. Utilizing sentiment analysis, this study attempts to identify Indonesian public sentiment and emotions toward the LSSR from a macro perspective. Several studies have examined public response toward government rules concerning COVID-19 prevention in Indonesia using sentiment analysis, for example, Djalante et al [11] and Raamkumar et al in Singapore [21]. However, both studies looked at responses on specific platforms and cannot be generalized to the larger population. Therefore, this study will offer a macro approach by mining and analyzing big data from more than 500 online news portals and social media platforms such as Twitter, Facebook, Instagram, and YouTube. This study is useful to gauge public responses in Indonesia in detail as guidance to improve existing policy.

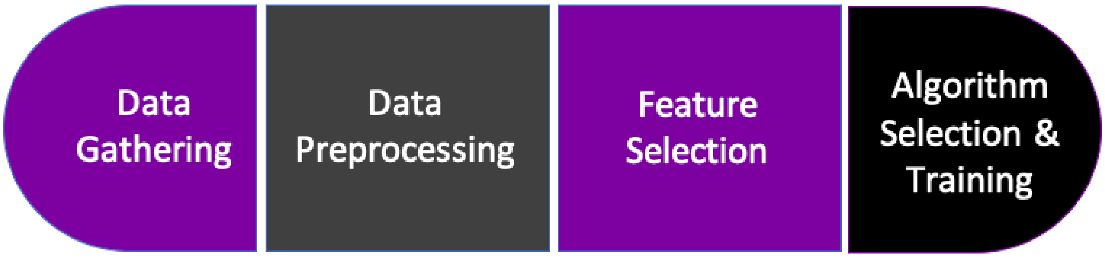
Methods

Study Design

This study employed sentiment analysis using a supervised machine learning approach to mine COVID-19-related posts on selected media platforms (Twitter, Facebook, Instagram, and YouTube). Machine learning was chosen as the method for classifying sentiment categories considering its effectiveness to accommodate a larger data set. The analysis was also performed on COVID-19 news published in more than 500 online news platforms recognized by the Indonesian Press Council. Online news and social media posts included in the analysis are those posted between March 31 and May 31, 2020. The whole process of supervised machine learning application is described in Figure 1.

The first analysis in this study involved sentiment analysis on data from both online news portals and social media platforms. The second analysis was an emotion analysis conducted on the Twitter platform to identify collective public emotion toward the LSSR. The whole analysis process on social media and online news portal involved natural language processing (NLP) on Apache SOLR. Details of the systematic analysis process are as outlined below.

Figure 1. Machine learning workflow.



Data Gathering

This study applied 2 different methods for data collection from each platform. The data from online news portals were collected using Perl Web Crawler, while real-time application programming interface (API) provided by Perl Programming Language, including Net::Twitter, Facebook::Graph, API::Instagram, and WebService::YouTube, was employed for each social media platform. Online news and social media posts obtained were then stored in the Apache SOLR database. Search profiling was accomplished using the terms "*Pembatasan Sosial Berskala Besar*" (large-scale social restrictions), "*Pembatasan Sosial*" (Social Restrictions), and "*PSBB*" (LSSR) as the main keywords to describe the government-implemented social restrictions in Indonesia. These keywords were deemed sufficient to capture public interest in the LSSR as there are no other synonyms in either Bahasa Indonesia or English that refer specifically to the government-implemented restrictions.

Data Preprocessing

Overview

Additionally, a data cleansing process was conducted. The supervised machine learning approach enabled the researchers to sort, select, and also update the data stored in the search engine to ensure that only data associated with the study context will be analyzed. The main purpose of data preprocessing is to clean the data set from all noise and outliers. The details of this phase are described in the following subsections.

Duplicate Text Filtering

This process was done to remove text duplication and ensure only original data were included for analysis. All duplications in social media data (mostly from retweets) and online news were removed to avoid redundancy.

Text Normalization

All usernames, stop words (eg, at, in, on, may, always), URLs, hash symbols (#), punctuation marks, and other nonalphanumeric characters were removed as those entities would have no influence in determining the value of sentiment. Besides, typos and multiple occurrences of certain characters within 1 word were normalized in this stage (eg, "SAAAAAAAAAD" becomes "SAD").

Case Folding

Case folding is a procedure that helps data normalization, in which all texts are standardized to lower case to make it easier for the system to retrieve information more effectively.

Feature Selection (Part-of-Speech Tagging)

The part-of-speech (POS) tagging system is a word processing tool in NLP that refers to the process of tagging text onto a corresponding POS based on its definition and its relation with other words [22]. Commonly, only words that are categorized as nouns, adjectives, verbs, and adverbs were tokenized and tagged by the POS tagger, as those words are considered important indicators of objectivities and opinions [23]. Some studies have been conducted to create an Indonesian Standardized Annotated Corpus by applying a probability statistic approach such as Hidden Markov Model [24] and Maximum Entropy [25]; nevertheless, none of them are accessible for general audiences [26]. Thus, in this study, it was decided that the Indonesian Language Dictionary (*Kamus Besar Bahasa Indonesia* or *KBBI*) would be used as the main lexical database.

Algorithm Selection and Training

The naïve Bayes classifier is a machine learning algorithm that is commonly utilized for text classification based on probabilistic calculations to predict future possibilities by looking at past patterns [27]. Although there is still no exact justification on which algorithm is best in classifying text, the use of naïve Bayes has been proven effective or even more accurate [28] and precise [23,29] than other algorithms in previous studies. Additionally, a recent study discovered naïve Bayes to be more accurate in determining COVID-19 sentiment categories compared with other classifiers such as logistic regression [30]. Therefore, it was decided that this algorithm will be used to classify the sentiment into 3 categories from the training data: positive, negative, and neutral. The use of naïve Bayes algorithm for this study can be described by the following equation:

where V_{MAP} is all examined categories; V_c is the sentiment category; $P(X_i|V_c)$ is the probability of the word i falling into category c; and $P(V_c)$ is the probability of V_c.

Large amounts of data tokenized into nouns, adjectives, verbs, and adverbs by POS tagging were stored within a training data set and trained using fivefold cross validation to ensure the accuracy of the sentiments. To further improve the accuracy, a system that allows admin intervention in the sentiment labeling process was developed and interannotator agreement was applied involving 3 annotators (from Astramaya) in the review process to ensure the validity of labeling.

In the review process, all unrelated contents were removed as they are not suitable for analysis. Further, this process enabled the researchers to fix the sentiment defined on the machine so that the sentiment tendency would fit to the context and demands of the analysis. After the selection process, we obtained 1,934,575 mentions of the search terms on both online news portals and social media platforms from March 31 to May 31, 2020. The data crawling process placed Twitter as the platform with the highest number of data, in which the keywords were mentioned 1,440,062 times. The use of Twitter for sentiment analysis research is indeed more popular than other platforms. Twitter is accessed by over 330 million users monthly and has 145 million daily active users [31]. The open nature of the microblogging service with numerous daily messages produced and generated has positioned it as the focal point of social media research and NLP [32].

Lexicon-Based Approach for Emotion Analysis

The emotion analysis in this study was performed using a lexicon-based, or specifically a dictionary-based approach, in which the Indonesian Language dictionary *Kamus Besar Bahasa Indonesia* (*KBBI*) was utilized as the lexical database. The categorization of emotion was derived from the basic emotion concept put forth by Plutchik [33], including joy, trust, fear, surprise, sadness, disgust, anger, and anticipation. There is a slight methodology difference in determining emotion categories for the data collected from both types of platforms. On social media, public sentiment and emotion were identified based on an analysis of the text in each post; 1 post or unit of analysis could represent 1 or more emotions. Emotion analysis with a lexicon-based approach enables the researcher to conduct pattern matching using the regular expression feature on the Perl application. In addition, emotion analysis was not carried out for online news platform as online news does not represent any form of public conversation trend. Thus, only public sentiment can be obtained on this platform, and not public emotion.

Data Visualization

Visualization is the final part of the NLP process, where the results of the analysis are translated into visuals to support the statistical data. To optimize the data presentation interface, all visual presentations in this study were delivered using Zend Framework (Linux Foundation).

Results

Data Demography

Search using the keywords "*Pembatasan Sosial*" and "*PSBB*" (LSSR) performed on both online news and social media platforms revealed 1,934,596 mentions of these in news sentences and social media posts. Twitter yielded the highest mentions (1,440,062/1,934,596, 74.44%), followed by online news (397,004/1,934,596, 20.52%), Instagram (50,754/1,934,596, 2.62%), Facebook (43,819/1,934,596, 2.27%), and YouTube (2957/1,934,596, 0.15%). Figure 2 presents the data distribution for each platform.

The data distribution showed that discussions related to LSSR on social media mainly took place on Twitter. This justified the use of Twitter to describe the overall sentiment and emotions toward the LSSR on social media. To identify popular topics discussed by the public, the NLP process helped to cluster discourses which were predominantly used and written in online news platforms via topic mapping. However, on Twitter, popular topics were identified based on top tweets during the study period. Figure 3 reports discourses and topic mapping for LSSR in online news platforms.

Figure 2. Data distribution for each platform.

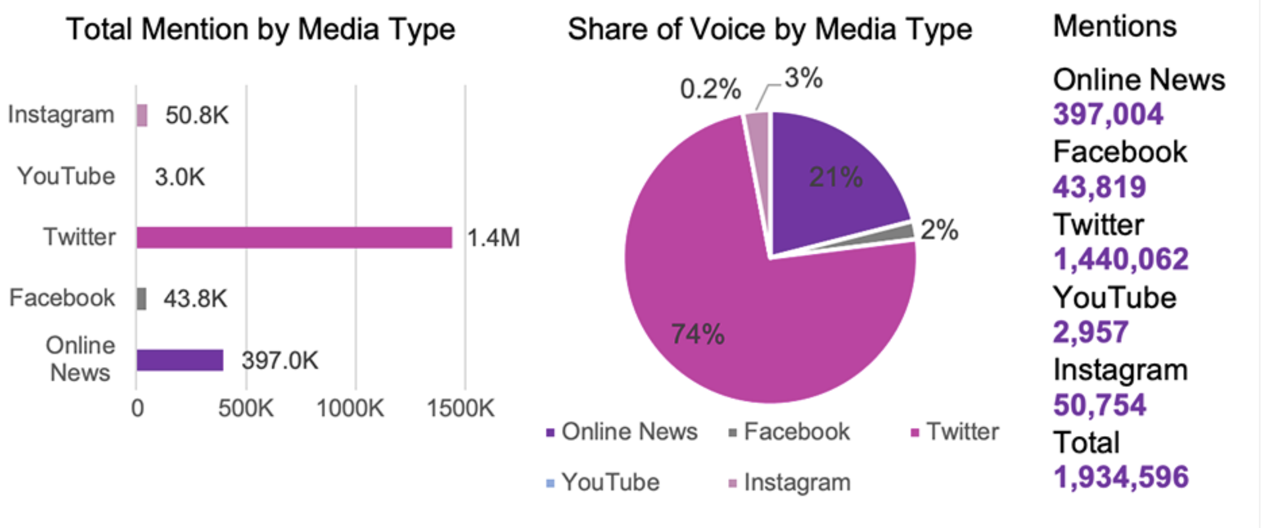
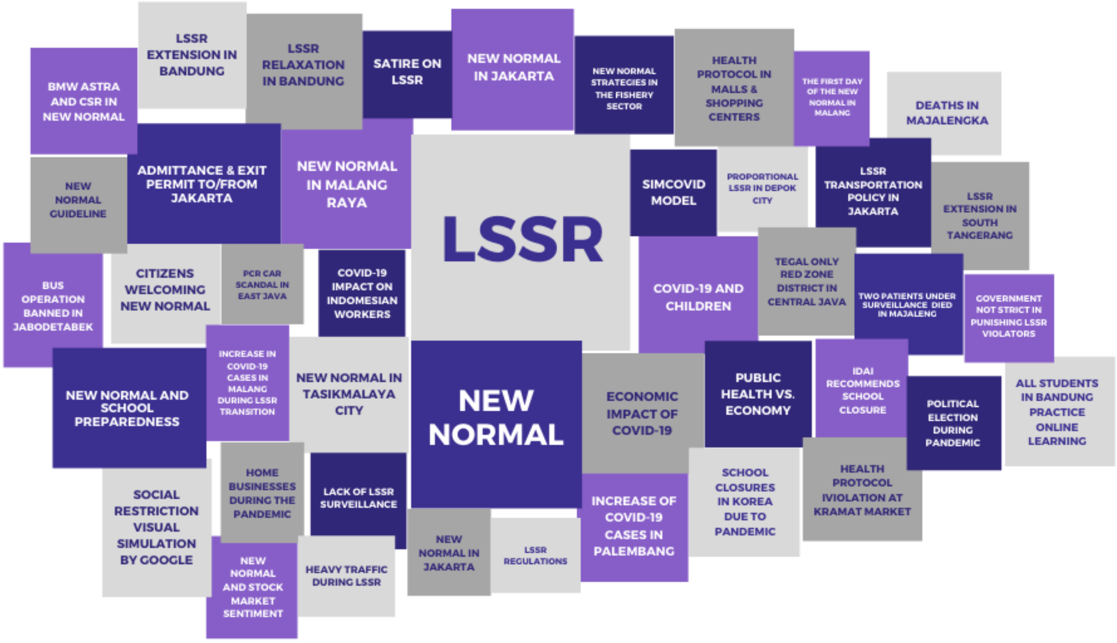


Figure 3. Topic mapping of online news platforms.

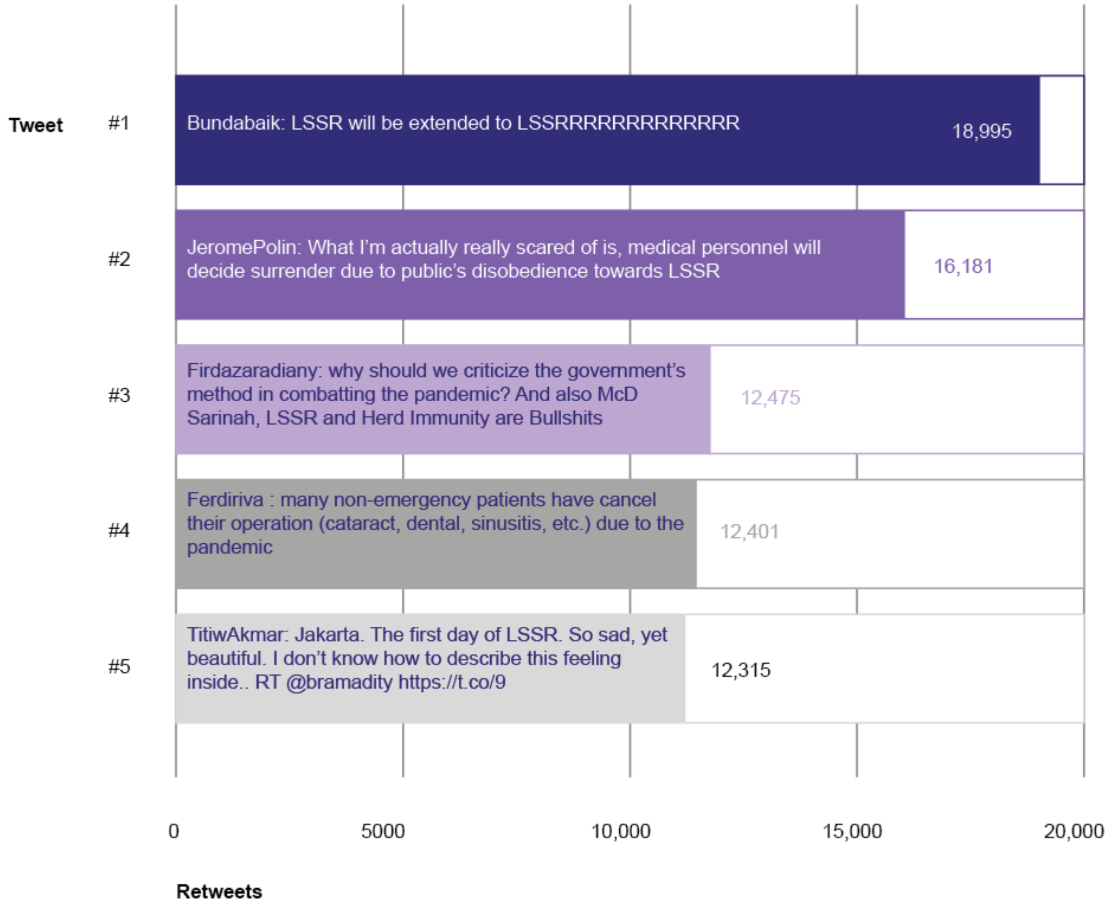

LSSR EXTENSION IN BANDUNG
LSSR RELAXATION IN BANDUNG
SATIRE ON LSSR
NEW NORMAL IN JAKARTA
NEW NORMAL STRATEGIES IN THE FISHERY SECTOR
HEALTH PROTOCOL IN MALLS & SHOPPING CENTERS
THE FIRST DAY OF THE NEW NORMAL IN MALANG
DEATHS IN MAJALENGKA
BMW ASTRA AND CSR IN NEW NORMAL
NEW NORMAL GUIDELINE
ADMITTANCE & EXIT PERMIT TO/FROM JAKARTA
NEW NORMAL IN MALANG RAYA
LSSR
SIMCOVID MODEL
PROPORTIONAL LSSR IN DEPOK CITY
LSSR TRANSPORTATION POLICY IN JAKARTA
LSSR EXTENSION IN SOUTH TANGERANG
BUS OPERATION BANNED IN JABODETABEK
CITIZENS WELCOMING NEW NORMAL
PCR CAR SCANDAL IN EAST JAVA
COVID-19 IMPACT ON INDOMESIAN WORKERS
COVID-19 AND CHILDREN
TEGAL ONLY RED ZONE DISTRICT IN CENTRAL JAVA
TWO PATIENTS UNDER SURVEILLANCE DIED IN MAJALENG
GOVERNMENT NOT STRICT IN PUNISHING LSSR VIOLATORS
NEW NORMAL AND SCHOOL PREPAREDNESS
INCREASE IN COVID-19 CASES IN MALANG DURING LSSR TRANSITION
NEW NORMAL IN TASIKMALAYA CITY
NEW NORMAL
ECONOMIC IMPACT OF COVID-19
PUBLIC HEALTH VS. ECONOMY
IDAI RECOMMENDS SCHOOL CLOSURE
POLITICAL ELECTION DURING PANDEMIC
ALL STUDENTS IN BANDUNG PRACTICE ONLINE LEARNING
SOCIAL RESTRICTION VISUAL SIMULATION BY GOOGLE
HOME BUSINESSES DURING THE PANDEMIC
LACK OF LSSR SURVEILLANCE
SCHOOL CLOSURES IN KOREA DUE TO PANDEMIC
HEALTH PROTOCOL IVIOLATION AT KRAMAT MARKET
INCREASE OF COVID-19 CASES IN PALEMBANG
NEW NORMAL AND STOCK MARKET SENTIMENT
HEAVY TRAFFIC DURING LSSR
NEW NORMAL IN JAKARTA
LSSR REGULATIONS

The most frequently discussed topics related to LSSR in Indonesia are LSSR policy, the new normal, and the impact of LSSR on the economic sector. The high level of conversation regarding the relaxation of LSSR in online news platforms represents the amplification of various responses from the public and media toward the government's swift plan to implement new norms.

The discussions about LSSR on both online news platforms and social media are mostly in relation to the new normal. In addition, news crawling found that a number of LSSR-related news are political. Many assumed that the LSSR relaxation in Indonesia was not based on careful consideration but was implemented to reduce the economic burden on the government. The term "political crisis" was used to describe what would happen if the government did not take any immediate action to ease the economic burden of its society. This assumption is supported by the high number of new reports highlighting the economic impact of LSSR on the nation. Government as the policy maker faced various negative reverberations about LSSR relaxation, as it was seen to be prioritizing the economy over public health.

In addition, to identify LSSR topics with the highest engagement level on Twitter, tweets with the highest number of retweets were ranked. Figure 4 describes 5 tweets with the highest engagement.

Figure 4. Most retweeted tweets.

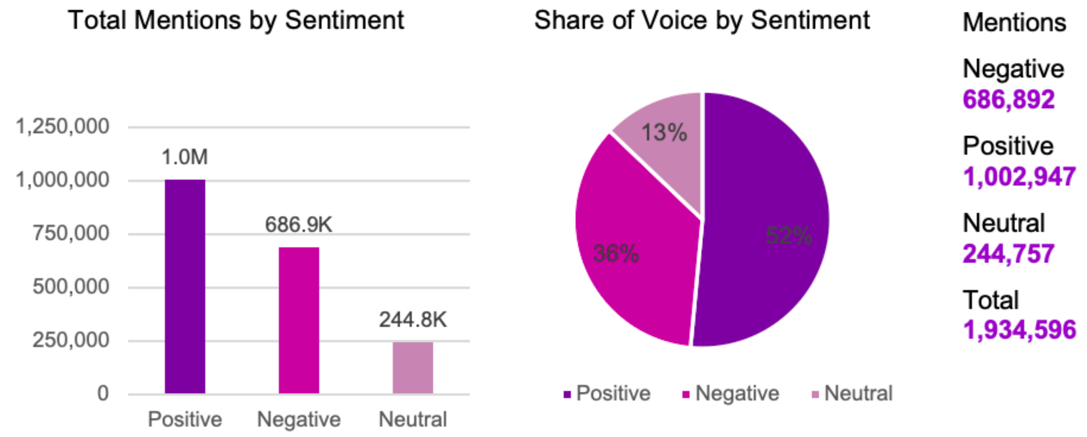


Overall, it can be assumed that the tweets represented public outburst and resentment toward 2 parties: the government that is considered indecisive to implement LSSR, and those among the public who enjoyed the LSSR relaxation by going out freely.

Public Sentiment Toward Large-Scale Social Restrictions

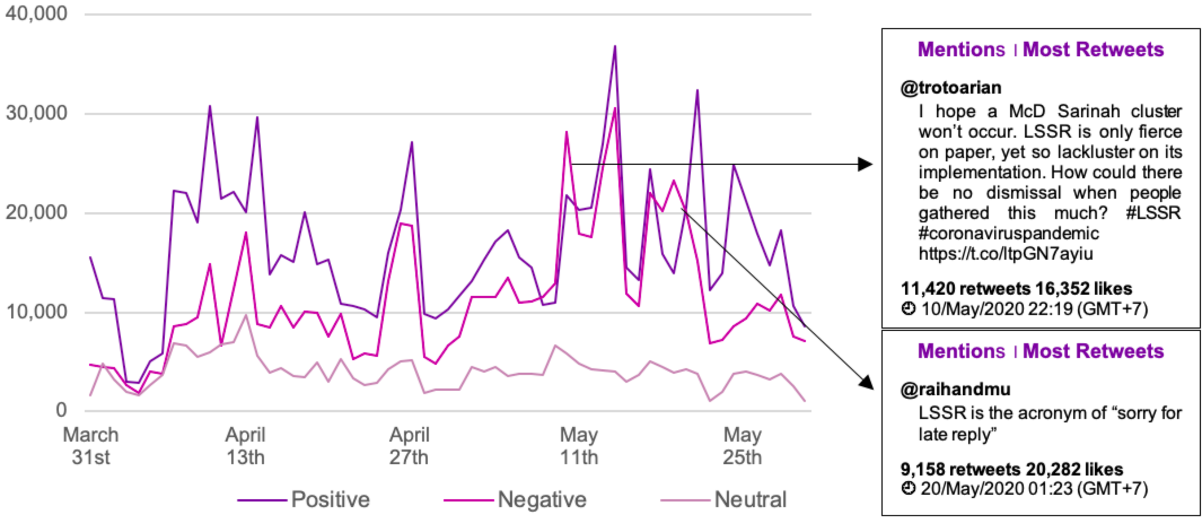
Sentiment analysis on online news platforms and social media during the period between March 31 and May 31, 2020 (n=1,934,596) showed that public sentiment toward the LSSR tended to be positive. Figure 5 presents the distribution of total mentions of positive, negative, and neutral sentiments in online news and social media platforms.

Figure 5. Total mentions by sentiment.



Based on the statistics in Figure 5, positive sentiment surpasses other sentiment categories (1,002,947/1,934,596, 51.84%) followed by negative (686,892/1,934,596, 35.51%) and neutral (244,757/1,934,596, 12.65%) sentiments. These data suggest that although society in general viewed the government's LSSR implementation and relaxation as indecisive, the majority still supported and regarded LSSR as the best solution to break the chain of COVID-19 transmission in the country. At the same time the negative sentiment on LSSR can be considered very high. It is not a good indicator of government performance, moreover, considering that the recovery phase of this health crisis will highly depend on public support and cooperation. Therefore, strategic and proactive efforts are needed to maximize the support from society. To deepen the understanding of public sentiment, a trend analysis was carried out as outlined in Figure 6.

Figure 6. Sentiment trends in all media types.



@trotoarian
I hope a McD Sarinah cluster won't occur. LSSR is only fierce on paper, yet so lackluster on its implementation. How could there be no dismissal when people gathered this much? #LSSR #coronaviruspandemic https://t.co/ltpGN7ayiu

11,420 retweets 16,352 likes
10/May/2020 22:19 (GMT+7)

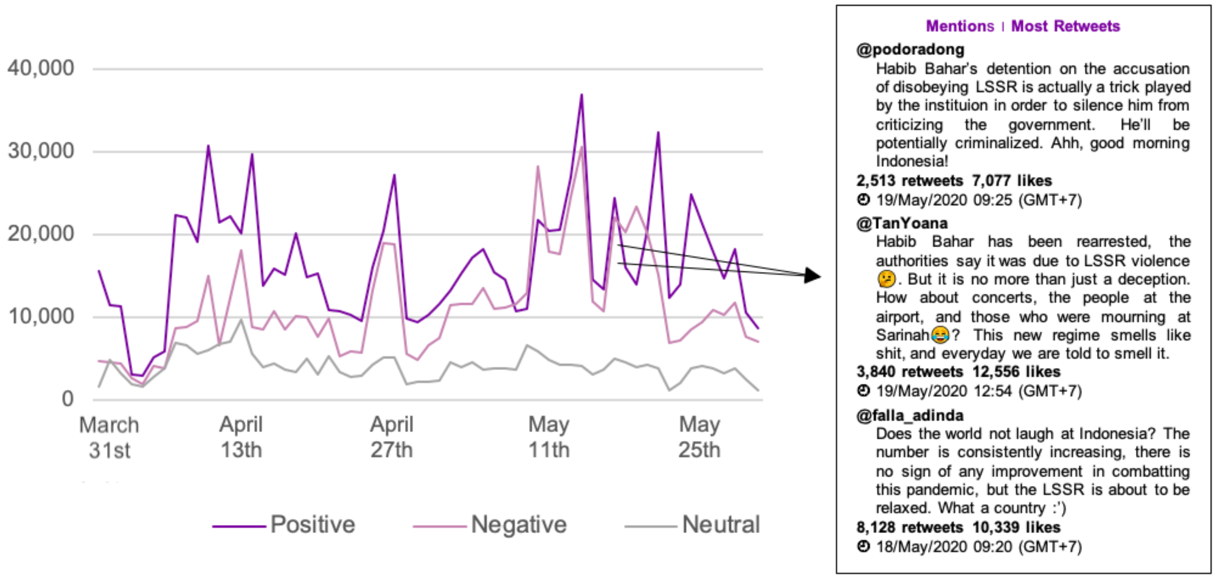
@raihandmu
LSSR is the acronym of "sorry for late reply"

9,158 retweets 20,282 likes
20/May/2020 01:23 (GMT+7)

Figure 6 presents the overall trend in sentiment for both online news and social media. It was observed that overall, positive sentiments were higher than negative and neutral sentiments. There were only 2 days where negative sentiments were higher than positive sentiments, precisely on May 11 and 21, 2020. Tweets with the highest engagement on May 11, 2020, highlighted the flaws in LSSR implementation, wherein the government lacked a firm response toward a mass gathering in Sarinah, Central Jakarta.

The positive and negative sentiments toward the LSSR reached the peak on the same day (ie, May 14, 2020), with recorded mentions of more than 30,000. Interestingly, both positive and negative opinions on that day were referring to a particular tweet, posted by @Jeromepolin, regretting the public's noncompliance to the LSSR. A detailed analysis concerning the discussion trend revealed that the negative sentiments referred mostly to the indecisive government and noncompliance toward the LSSR. A similar pattern of both positive and negative sentiments with high mentions was also recorded on May 17, 2020 (Figure 7).

Figure 7. Sentiment trends on May 17, 2020.



@podoradong
Habib Bahar's detention on the accusation of disobeying LSSR is actually a trick played by the instituion in order to silence him from criticizing the government. He'll be potentially criminalized. Ahh, good morning Indonesia!
2,513 retweets 7,077 likes
🕘 19/May/2020 09:25 (GMT+7)

@TanYoana
Habib Bahar has been rearrested, the authorities say it was due to LSSR violence 🤔. But it is no more than just a deception. How about concerts, the people at the airport, and those who were mourning at Sarinah😂? This new regime smells like shit, and everyday we are told to smell it.
3,840 retweets 12,556 likes
🕘 19/May/2020 12:54 (GMT+7)

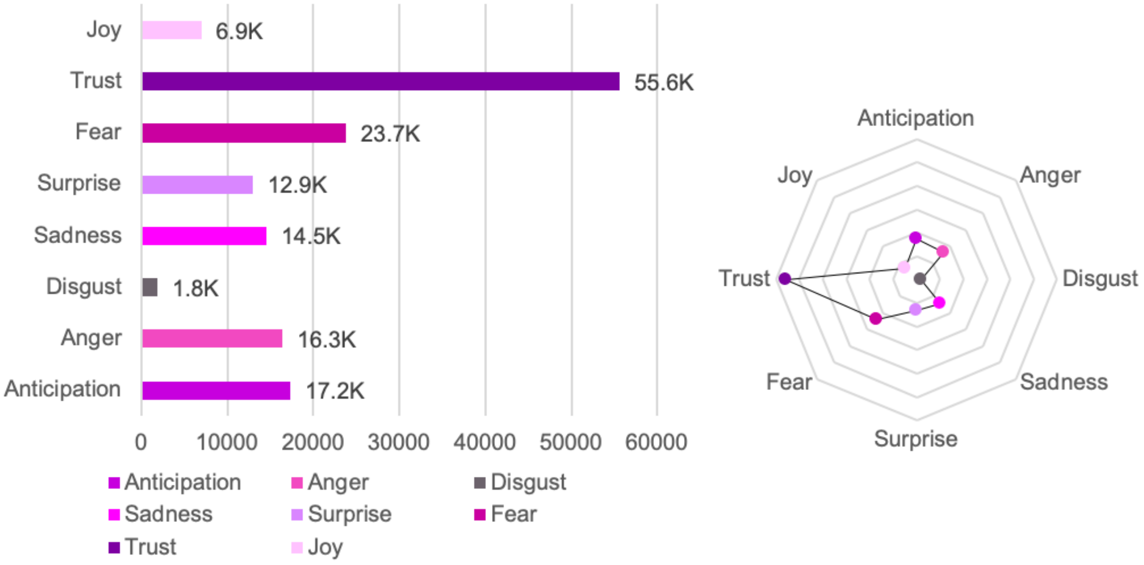
@falla_adinda
Does the world not laugh at Indonesia? The number is consistently increasing, there is no sign of any improvement in combatting this pandemic, but the LSSR is about to be relaxed. What a country :')
8,128 retweets 10,339 likes
🕘 18/May/2020 09:20 (GMT+7)

Similar to the sentiment trends on May 14, 2020, the topic with the highest mentions on May 17, 2020, also referred to skepticism toward the government's implementation of the LSSR. In addition to this, some parties were disappointed with the government's decision to relax the LSSR when positive cases of COVID-19 in Indonesia were still on the rise.

Public Emotions Toward LSSR

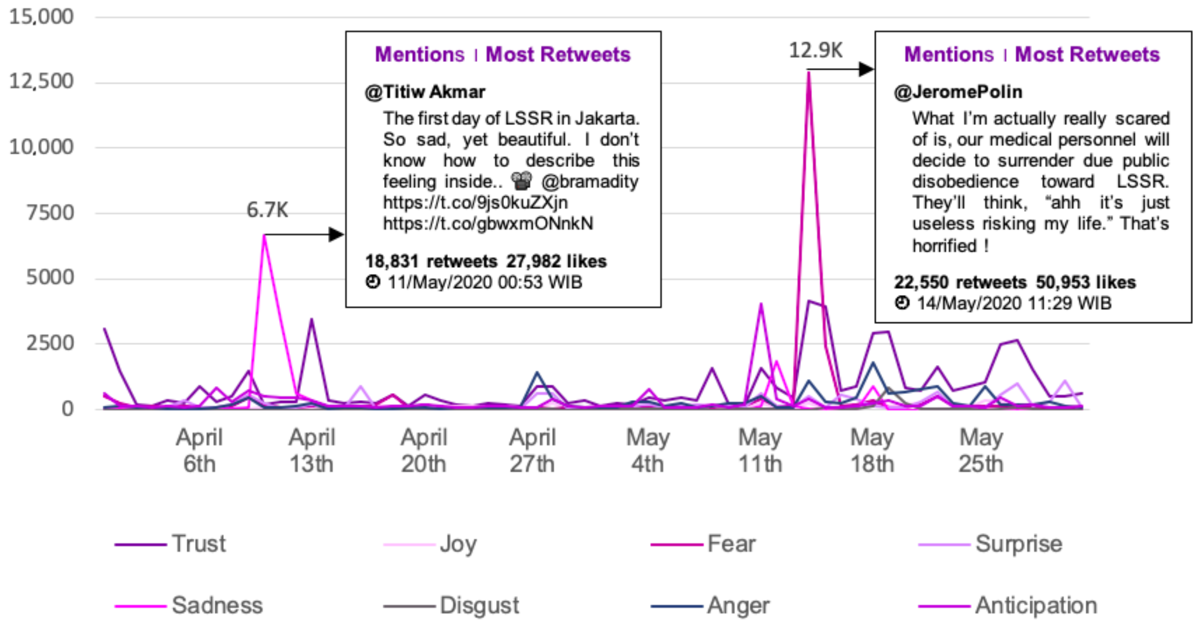
Public emotion analysis using a lexicon-based approach and quote extraction was applied to explore collective public emotion regarding the LSSR. As discussed in the "Methods" section, public emotion was categorized based on the basic emotion concept by Plutchik [33], including joy, trust, surprise, fear, disgust, sadness, anger, and anticipation. Furthermore, this analysis can be used as a parameter to identify how far the government and the LSSR garnered public trust. Public emotions toward the LSSR are shown in Figure 8.

Figure 8. Emotion analysis.



The emotion analysis concluded that although there were many negative opinions toward the government and the LSSR, the public still believed that the LSSR was the best method to break the chain of COVID-19 transmission. The public expected that all members of society would accept and comply with this policy in order to achieve overall success. At the same time, the public still had expectations that the government would immediately overcome the issue by implementing firmer control. Although the overall public emotion was still "trustful," interesting trends were found in 2 other emotion categories, namely, "fear" and "sadness." Figure 9 depicts public emotion trends during the study period.

Figure 9. Trend of public emotions.



Mentions | Most Retweets

@Titiw Akmar

The first day of LSSR in Jakarta. So sad, yet beautiful. I don't know how to describe this feeling inside.. 📽 @bramadity https://t.co/9js0kuZXjn https://t.co/gbwxmONnkN

18,831 retweets 27,982 likes
🕘 11/May/2020 00:53 WIB

Mentions | Most Retweets

@JeromePolin

What I'm actually really scared of is, our medical personnel will decide to surrender due public disobedience toward LSSR. They'll think, "ahh it's just useless risking my life." That's horrified !

22,550 retweets 50,953 likes
🕘 14/May/2020 11:29 WIB

The "sadness" and "fear" emotions showed the most outstanding patterns compared with other emotion categories and were found to reach their peak on April 11, 2020, and May 14, 2020, respectively. On April 11, 2020, the most retweeted post suggested public sadness toward the implications of the LSSR. Moreover, on that date, the number of deaths continued to surpass the total recovered cases in the country. The same trend occurred on 22 May, 2020, which also indicated "fear" of public disobedience toward the LSSR.

Discussion

Principal Findings

This study found Twitter as the most dominant platform used by the public when discussing COVID-19 and LSSR. Public sentiment toward the LSSR was largely positive with expressions of support for the government intervention in combating COVID-19. Even so, negative sentiments were also quite high, especially concerning government's indecisiveness and public defiance of the LSSR. The emotion analysis revealed that Indonesians were mostly trustful with notable peaks in sadness and fear corresponding to increased COVID-19 death rates.

A closer examination of tweets with the highest level of engagement showed that a majority of the public were dissatisfied with the government and parties that did not obey LSSR regulations. This is supported by the trends observed in the sentiment analysis. Criticisms of government's indecisiveness and LSSR relaxation dominated the discussion trends regarding LSSR during the study period. Even though the overall public sentiment was positive, the negative and neutral sentiments were still considered high. The fact that almost half of the public did not show positive attitudes toward the government and LSSR implementation can be worrying, particularly in a global health crisis of this magnitude. A neighboring country, Vietnam, for example, has proven that effective cooperation between the society and the government can successfully allow the country to control the spread of COVID-19 cases. Compliance of Vietnamese to their government instructions such as wearing masks and implementing self-quarantine [34] has made their country one of the victorious in managing the pandemic with less than 60 deaths as of June 2021 [35].

The results also show that the Indonesian government's inconsistent stance in implementing policies has contributed to the emergence of public emotions such as fear and sadness. The Indonesian media may have played a role in this, as concerns about socioeconomic uncertainty, LSSR, social distancing violations, and economic downturn in all sectors were widely reported in the media from March to early April 2020 [36]. In the earlier phase of the pandemic, negative media reports on issues such as insufficient government response, patient care, and burial of the deceased also caused agitation among the Indonesian public [37]. A previous study on COVID-19 news in Indonesia also suggested that the distortion of news by both online and mainstream media has caused an increase in public fear and anxiety [38].

Analysis of sentiments surrounding COVID-19 has produced similar results in communities across the globe. The health crisis has sparked concerns on the spread of the virus and its containment [39], economic survival [40], and physical and mental well-being [41]. The uncertainty surrounding the pandemic has resulted in expressions of mixed emotions in the public domain. An analysis of Twitter discussions in the early stages of the pandemic found dominant emotions of trust, anger, and fear surrounding COVID-19 cases and deaths [39]. A study in Italy observed a similar pattern; feelings of anger and fear emerged along with trust, solidarity, and hope [42], while in Spain sentiments surrounding the pandemic were that of disgust, fear, anger, and sadness [43].

It is important to note that while these sentiments and emotions were dominant at the beginning of the pandemic in early 2020, there has been much polarity and change in the sentiments since then [44]. COVID-19 should be a lesson for governments to affirm the consequential role of strategic public policy and acknowledge the importance of clear, consistent, and well-targeted public communications in overcoming crisis.

While public trust is still high, the government must be responsive to the needs of its society, considering their outlook as the basis for formulating policies and revising those deemed incompatible with public needs. Economic struggles were one of the biggest drivers of LSSR noncompliance in Indonesia. The large number of violations found in big cities such as Jakarta, Bogor, Depok, Tangerang, and Bekasi [45] is consistent with the findings of Wasdani and Prasad [46], who reported that the implementation of social distancing among the urban poor is very difficult. Therefore, strategic public policy should be introduced to maintain economic resilience, for instance, allocating budget for economic stimulus packages as done by the Malaysian government, which distributed MYR 260 billion (US $61 billion) to its people in the lower- and medium-income groups [47].

Even so, robust public policy alone may not be sufficient to curb negative emotions surrounding the pandemic. Mohamad et al [48] have highlighted that the delivery of clear, consistent, and credible information is key to control and mitigate the disease. Misinformation resulting from the communication of policies greatly affects public perceptions and trust during the pandemic [49]. To ensure successful communication to the public, authorities must recognize the important role of mass media in delivering quality information, as evidenced by previous studies [43].

Strength and Limitations

The strength of this study is in the variety of platforms analyzed. To obtain more diverse data, analyzing various types of social media and online news platforms with relatively large data sets makes it possible to synthesize diverse results compared with previous studies. Previous studies involving online media often only analyzed a single type of platform such as Twitter [49,50], Sina Weibo, the Chinese version of Twitter [51,52], Facebook [21], and online news portals [53,54]. By contrast, this study has a relatively larger data set collected by optimizing web-crawling functions on more than 500 online news portals and large amounts of social media data.

However, there were some limitations to this study. The large amount of data analyzed on social media platforms made it difficult for the researchers to determine what topics were dominant on social media, particularly Twitter. The text-based approach used on Twitter requires a relatively longer issue clustering process. Accordingly, the topic mapping could only be used to analyze the data gathered from online news platforms, by applying a clustering method to identify topics that are similar to each other, assisted by paragraph segmentation to deepen researcher understanding of the news context. In addition, this study did not conduct emotion analysis on the comments that appeared in online news. Future research is recommended to explore public emotions by analyzing comments written in response to news items.

Research Implications

The results of this study may assist policy makers in being active observers of public responses in online media, especially on social media. Public communications by the government in this new media era must transform from being mere publicity to observing and analyzing public sentiment.

Social listening is becoming an important way to gauge public opinion and response. Public confusion surrounding the LSSR is a sign that the government must develop a comprehensive communication protocol to emphasize the importance of LSSR in eradicating the spread of COVID-19 in Indonesia.

Conclusions

Overall, this study disclosed Twitter as the most popular platform used by Indonesian public in conveying their thoughts regarding LSSRs. The analysis performed on the major keywords that appeared on social media and online news portals revealed positive sentiment toward the LSSR. However, there were concerns on emerging conversations regarding public skepticism toward effective implementation of the LSSR. In addition, the emotion analysis concluded that a majority of the public still believe LSSR as the best method to effectively slow the spread of COVID-19. This explains why the public still hold high hopes and expectations toward the government as the policy maker in keeping them safe in a health crisis such as this.

Acknowledgments

The completion of this paper could have not been done without outstanding support from our partner, Astramaya, in running the data mining procedure. Therefore, we express our sincere appreciation to and acknowledge the Astramaya team for their contribution. It was an honor to work with the biggest and most reputable data analytics company in Indonesia.

Conflicts of Interest

None declared.

References

1. Askitas N, Tatsiramos K, Verheyden B. Estimating worldwide effects of non-pharmaceutical interventions on COVID-19 incidence and population mobility patterns using a multiple-event study. Sci Rep 2021 Jan 21;11(1):1972 [FREE Full text] [doi: 10.1038/s41598-021-81442-x] [Medline: 33479325]

2. Mehta RM, Mehta R, Balaji AL, Mehta H. Perceptions on COVID19: A ground-level analysis to guide public policy. Lung India 2020;37(3):282-283 [FREE Full text] [doi: 10.4103/lungindia.lungindia_190_20] [Medline: 32367858]
3. Punn N, Sonbhadra S, Agarwal S. Monitoring Covid-19 Social Distancing with Person Detection and Tracking via Fine-tuned YOLO v3 and Deepsort Techniques. arXiv 2021 [FREE Full text]
4. Jarvis CI, Van Zandvoort K, Gimma A, Prem K, CMMID COVID-19 working group, Klepac P, et al. Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK. BMC Med 2020 May 07;18(1):124 [FREE Full text] [doi: 10.1186/s12916-020-01597-8] [Medline: 32375776]
5. Vogel L. How long will social distancing take to work? Experts weigh in on Canada's COVID-19 response. CMAJ 2020 Apr 06;192(14):E382-E383 [FREE Full text] [doi: 10.1503/cmaj.1095857] [Medline: 32392531]
6. Wong J, Koh WC, Alikhan MF, Abdul Aziz ABZ, Naing L. Responding to COVID-19 in Brunei Darussalam: Lessons for small countries. J Glob Health 2020 Jun;10(1):010363 [FREE Full text] [doi: 10.7189/jogh.10.010363] [Medline: 32566154]
7. Hsieh C, Lin C, Wang WYC, Pauleen DJ, Chen JV. The Outcome and Implications of Public Precautionary Measures in Taiwan-Declining Respiratory Disease Cases in the COVID-19 Pandemic. Int J Environ Res Public Health 2020 Jul 06;17(13):4877 [FREE Full text] [doi: 10.3390/ijerph17134877] [Medline: 32640752]
8. Ng CFS, Seposo XT, Moi ML, Tajudin MABA, Madaniyazi L, Sahani M. Characteristics of COVID-19 epidemic and control measures to curb transmission in Malaysia. Int J Infect Dis 2020 Dec;101:409-411 [FREE Full text] [doi: 10.1016/j.ijid.2020.10.027] [Medline: 33075527]
9. Abideen AZ, Mohamad FB, Hassan MR. Mitigation strategies to fight the COVID-19 pandemic—present, future and beyond. JHR 2020 Aug 26;34(6):547-562. [doi: 10.1108/JHR-04-2020-0109]
10. Setiati S, Azwar MK. COVID-19 and Indonesia. Acta Med Indones 2020 Jan;52(1):84-89 [FREE Full text] [Medline: 32291377]
11. Djalante R, Lassa J, Setiamarga D, Sudjatma A, Indrawan M, Haryanto B, et al. Review and analysis of current responses to COVID-19 in Indonesia: Period of January to March 2020. Prog Disaster Sci 2020 Apr;6:100091 [FREE Full text] [doi: 10.1016/j.pdisas.2020.100091] [Medline: 34171011]
12. Channel News Asia. Channel News Asia. URL: https://infographics.channelnewsasia.com/covid-19/map.html [accessed 2020-06-03]
13. INDEF: Indonesia's Large-Scale Social Restrictions Is The Worst Among Neighboring Country. Detik News. URL: https://news.detik.com/berita/d-5021739/indef-hasil-psbb-indonesia-paling-buruk-dibanding-negara-negara-tetangga/1 [accessed 2020-06-03]
14. McFadden SM, Malik AA, Aguolu OG, Willebrand KS, Omer SB. Perceptions of the adult US population regarding the novel coronavirus outbreak. PLoS One 2020;15(4):e0231808 [FREE Full text] [doi: 10.1371/journal.pone.0231808] [Medline: 32302370]
15. Mahmood S, Hussain T, Mahmood F, Ahmad M, Majeed A, Beg BM, et al. Attitude, Perception, and Knowledge of COVID-19 Among General Public in Pakistan. Front Public Health 2020;8:602434 [FREE Full text] [doi: 10.3389/fpubh.2020.602434] [Medline: 33363092]
16. Azlan AA, Hamzah MR, Sern TJ, Ayub SH, Mohamad E. Public knowledge, attitudes and practices towards COVID-19: A cross-sectional study in Malaysia. PLoS One 2020;15(5):e0233668 [FREE Full text] [doi: 10.1371/journal.pone.0233668] [Medline: 32437434]
17. Maheshwari S, Gupta P, Sinha R, Rawat P. Knowledge, attitude, and practice towards coronavirus disease 2019 (COVID-19) among medical students: A cross-sectional study. J Acute Dis 2020;9(3):100. [doi: 10.4103/2221-6189.283886]
18. Khasawneh AI, Humeidan AA, Alsulaiman JW, Bloukh S, Ramadan M, Al-Shatanawi TN, et al. Medical Students and COVID-19: Knowledge, Attitudes, and Precautionary Measures. A Descriptive Study From Jordan. Front Public Health 2020;8:253 [FREE Full text] [doi: 10.3389/fpubh.2020.00253] [Medline: 32574313]
19. Al-Hanawi MK, Angawi K, Alshareef N, Qattan AMN, Helmy HZ, Abudawood Y, et al. Knowledge, Attitude and Practice Toward COVID-19 Among the Public in the Kingdom of Saudi Arabia: A Cross-Sectional Study. Front Public Health 2020;8:217 [FREE Full text] [doi: 10.3389/fpubh.2020.00217] [Medline: 32574300]
20. Waitzberg R, Davidovitch N, Leibner G, Penn N, Brammli-Greenberg S. Israel's response to the COVID-19 pandemic: tailoring measures for vulnerable cultural minority populations. Int J Equity Health 2020 May 19;19(1):71 [FREE Full text] [doi: 10.1186/s12939-020-01191-7] [Medline: 32429995]
21. Sesagiri Raamkumar A, Tan SG, Wee HL. Measuring the Outreach Efforts of Public Health Authorities and the Public Response on Facebook During the COVID-19 Pandemic in Early 2020: Cross-Country Comparison. J Med Internet Res 2020 May 19;22(5):e19334 [FREE Full text] [doi: 10.2196/19334] [Medline: 32401219]
22. Alfred R, Mujat A, Obit JH. A ruled-based part of speech (RPOS) tagger for Malay text articles. 2013 Presented at: Proceedings of the Asian Conference on Intelligent Information and Database Systems; Mar; Kuala Lumpur, Malaysia; March 18-20, 2013; Kuala Lumpur, Malaysia p. 18-20. [doi: 10.1007/978-3-642-36543-0_6]
23. Vidya NA, Budi I, Fanany MI. Twitter sentiment to analyze net brand reputation of mobile phone providers. Amsterdam, The Netherlands: Elsevier; 2015 Presented at: Proceedings of the 3rd Information System International Conference; November 2-4, 2015; Surabaya, Indonesia p. 519-526. [doi: 10.1016/j.procs.2015.12.159]

24. Wicaksono AF, Purwarianti A. HMM based part-of-speech tagger for Bahasa Indonesia. 2010 Presented at: Proceedings of 4th International MALINDO (Malay and Indonesian Language) Workshop; August 2, 2010; Depok, Indonesia URL: https://www.researchgate.net/publication/209387036_HMM_Based_Part-of-Speech_Tagger_for_Bahasa_Indonesia
25. Nurwidyantoro A, Winarko E. Parallelization of maximum entropy POS tagging for Bahasa Indonesia with MapReduce. International Journal of Computer Science Issues 2012;9(4):1 [FREE Full text]
26. Muljono M, Afini U, Supriyanto C, Nugroho RA. The Development of Indonesian POS Tagging System for Computer-aided Independent Language Learning. Int. J. Emerg. Technol. Learn 2017 Nov 16;12(11):138. [doi: 10.3991/ijet.v12i11.7383]
27. Rozi IF, Hamdana EN, Iqbal Alfahmi MB. Development of twitter sentiment analysis using naïve bayes classifier method (case study of SAMSAT in Malang city). Jurnal Informatika Polinema 2018 Feb 01;4(2):149-154. [doi: 10.33795/jip.v4i2.164]
28. Sailunaz K, Alhajj R. Emotion and sentiment analysis from Twitter text. Journal of Computational Science 2019 Sep;36:101003. [doi: 10.1016/j.jocs.2019.05.009]
29. Le B, Nguyen N. Twitter sentiment analysis using machine learning techniques. In: Le Thi H, Nguyen N, Do T, editors. Advanced Computational Methods for Knowledge Engineering. Advances in Intelligent Systems and Computing. Cham, Switzerland: Springer; 2015:279-289.
30. Samuel J, Ali GGMN, Rahman MM, Esawi E, Samuel Y. COVID-19 public sentiment insights and machine learning for Tweets classification. Information 2020 Jun 11;11(6):314. [doi: 10.3390/info11060314]
31. Lin Y. 10 Twitter Statistics Every Marketer Should Know in 2020 [Infographic]. Obelro. 2021 May 30. URL: https://www.oberlo.com/blog/twitter-statistics [accessed 2020-06-04]
32. Stojanovski D, Strezoski G, Madjarov G, Dimitrovski I, Chorbev I. Deep neural network architecture for sentiment analysis and emotion identification of Twitter messages. Multimed Tools Appl 2018 Jun 18;77(24):32213-32242. [doi: 10.1007/s11042-018-6168-1]
33. Plutchik R. The Nature of Emotions. Am. Sci 2001;89(4):344. [doi: 10.1511/2001.4.344]
34. La V, Pham T, Ho M, Nguyen M, P. Nguyen K, Vuong T, et al. Policy Response, Social Media and Science Journalism for the Sustainability of the Public Health System Amid the COVID-19 Outbreak: The Vietnam Lessons. Sustainability 2020 Apr 07;12(7):2931-2957. [doi: 10.3390/su12072931]
35. Worldometer. URL: https://www.worldometers.info/coronavirus/ [accessed 2021-06-11]
36. Gandasari D, Dwidienawati D. Content analysis of social and economic issues in Indonesia during the COVID-19 pandemic. Heliyon 2020 Nov;6(11):e05599 [FREE Full text] [doi: 10.1016/j.heliyon.2020.e05599] [Medline: 33305034]
37. Widati S, Soedirham O, Wee EH. Social construction of COVID-19 destigmatization through Indonesia's online mass media. International Journal of Current Research 2021;13(1):15870-15874. [doi: 10.24941/ijcr.40103.01.2021]
38. Sulistiadi W, Rahayu S, Harmani N. Handling of Public Stigma on COVID-19 in Indonesian Society. Jurnal Kesehatan Masyarakat Nasional (National Public Health Journal) 2020 Jul 27(Special Issue 1):70-76. [doi: 10.21109/kesmas.v15i2.3909]
39. Xue J, Chen J, Hu R, Chen C, Zheng C, Su Y, et al. Twitter Discussions and Emotions About the COVID-19 Pandemic: Machine Learning Approach. J Med Internet Res 2020 Nov 25;22(11):e20550 [FREE Full text] [doi: 10.2196/20550] [Medline: 33119535]
40. El Keshky MES, Basyouni SS, Al Sabban AM. Getting Through COVID-19: The Pandemic's Impact on the Psychology of Sustainability, Quality of Life, and the Global Economy - A Systematic Review. Front Psychol 2020;11:585897 [FREE Full text] [doi: 10.3389/fpsyg.2020.585897] [Medline: 33281683]
41. Wang C, Tee M, Roy AE, Fardin MA, Srichokchatchawan W, Habib HA, et al. The impact of COVID-19 pandemic on physical and mental health of Asians: A study of seven middle-income countries in Asia. PLoS One 2021;16(2):e0246824 [FREE Full text] [doi: 10.1371/journal.pone.0246824] [Medline: 33571297]
42. Stella M, Restocchi V, De Deyne S. #lockdown: Network-Enhanced Emotional Profiling in the Time of COVID-19. BDCC 2020 Jun 16;4(2):14. [doi: 10.3390/bdcc4020014]
43. de Las Heras-Pedrosa C, Sánchez-Núñez P, Peláez JI. Sentiment Analysis and Emotion Understanding during the COVID-19 Pandemic in Spain and Its Impact on Digital Ecosystems. Int J Environ Res Public Health 2020 Jul 31;17(15):5542 [FREE Full text] [doi: 10.3390/ijerph17155542] [Medline: 32751866]
44. Wicke P, Bolognesi MM. Covid-19 Discourse on Twitter: How the Topics, Sentiments, Subjectivity, and Figurative Frames Changed Over Time. Front. Commun 2021 Mar 16;6:651997. [doi: 10.3389/fcomm.2021.651997]
45. Taufan SA. Large-scale Social Restrictions Violation in Jadetabek has reached 70 thousand cases. Jawa Pos. URL: https://www.jawapos.com/jabodetabek/21/05/2020/pelanggaran-psbb-di-jadetabek-tembus-70-ribu/ [accessed 2020-06-06]
46. Wasdani KP, Prasad A. The impossibility of social distancing among the urban poor: the case of an Indian slum in the times of COVID-19. Local Environment 2020 Apr 21;25(5):414-418. [doi: 10.1080/13549839.2020.1754375]
47. Aziz NA, Othman J, Lugova H, Suleiman A. Malaysia's approach in handling COVID-19 onslaught: Report on the Movement Control Order (MCO) and targeted screening to reduce community infection rate and impact on public health and economy. J Infect Public Health 2020 Dec;13(12):1823-1829 [FREE Full text] [doi: 10.1016/j.jiph.2020.08.007] [Medline: 32896496]
48. Mohamad E, Tham JS, Ayub SH, Hamzah MR, Hashim H, Azlan AA. Relationship Between COVID-19 Information Sources and Attitudes in Battling the Pandemic Among the Malaysian Public: Cross-Sectional Survey Study. J Med Internet Res 2020 Nov 12;22(11):e23922 [FREE Full text] [doi: 10.2196/23922] [Medline: 33151897]

49. Sell TK, Hosangadi D, Trotochaud M. Misinformation and the US Ebola communication crisis: analyzing the veracity and content of social media messages related to a fear-inducing infectious disease outbreak. BMC Public Health 2020 May 07;20(1):550 [FREE Full text] [doi: 10.1186/s12889-020-08697-3] [Medline: 32375715]
50. 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]
51. 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]
52. 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]
53. Liu Q, Zheng Z, Zheng J, Chen Q, Liu G, Chen S, et al. Health Communication Through News Media During the Early Stage of the COVID-19 Outbreak in China: Digital Topic Modeling Approach. J Med Internet Res 2020 Apr 28;22(4):e19118 [FREE Full text] [doi: 10.2196/19118] [Medline: 32302966]
54. Rovetta A, Bhagavathula AS. COVID-19-Related Web Search Behaviors and Infodemic Attitudes in Italy: Infodemiological Study. JMIR Public Health Surveill 2020 May 05;6(2):e19374 [FREE Full text] [doi: 10.2196/19374] [Medline: 32338613]

Abbreviations

API: application programming interface
KBBI: *Kamus Besar Bahasa Indonesia*
LSSR: large-scale social restrictions
NLP: natural language processing
POS: part-of-speech

Edited by C Basch; submitted 26.02.21; peer-reviewed by K Usher, K Zeraatkar, A Rovetta; comments to author 26.04.21; revised version received 16.05.21; accepted 15.06.21; published 09.08.21.

Please cite as:
Tri Sakti AM, Mohamad E, Azlan AA
Mining of Opinions on COVID-19 Large-Scale Social Restrictions in Indonesia: Public Sentiment and Emotion Analysis on Online Media
J Med Internet Res 2021;23(8):e28249
URL: https://www.jmir.org/2021/8/e28249
doi:10.2196/28249
PMID:34280116

©Andi Muhammad Tri Sakti, Emma Mohamad, Arina Anis Azlan. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 09.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Forecasting the COVID-19 Epidemic by Integrating Symptom Search Behavior Into Predictive Models: Infoveillance Study

Alessandro Rabiolo[1*], MD; Eugenio Alladio[2*], PhD; Esteban Morales[3], MSc; Andrew Ian McNaught[1,4], MD; Francesco Bandello[5], MD; Abdelmonem A Afifi[6], PhD; Alessandro Marchese[5], MD

[1]Department of Ophthalmology, Gloucestershire Hospitals NHS Foundation Trust, Cheltenham, United Kingdom
[2]Department of Chemistry, University of Turin, Turin, Italy
[3]Jules Stein Eye Institute, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, United States
[4]School of Health Professions, Faculty of Health, University of Plymouth, Plymouth, United Kingdom
[5]Department of Ophthalmology, Vita-Salute University, IRCCS Ospedale San Raffaele Scientific Institute, Milan, Italy
[6]Department of Biostatistics, Fielding School of Public Health, University of California Los Angeles, Los Angeles, CA, United States
[*]these authors contributed equally

Corresponding Author:
Alessandro Rabiolo, MD
Department of Ophthalmology
Gloucestershire Hospitals NHS Foundation Trust
Sandford Rd
Cheltenham,
United Kingdom
Phone: 44 300 422 2527
Email: rabiolo.alessandro@gmail.com

Abstract

Background: Previous studies have suggested associations between trends of web searches and COVID-19 traditional metrics. It remains unclear whether models incorporating trends of digital searches lead to better predictions.

Objective: The aim of this study is to investigate the relationship between Google Trends searches of symptoms associated with COVID-19 and confirmed COVID-19 cases and deaths. We aim to develop predictive models to forecast the COVID-19 epidemic based on a combination of Google Trends searches of symptoms and conventional COVID-19 metrics.

Methods: An open-access web application was developed to evaluate Google Trends and traditional COVID-19 metrics via an interactive framework based on principal component analysis (PCA) and time series modeling. The application facilitates the analysis of symptom search behavior associated with COVID-19 disease in 188 countries. In this study, we selected the data of nine countries as case studies to represent all continents. PCA was used to perform data dimensionality reduction, and three different time series models (error, trend, seasonality; autoregressive integrated moving average; and feed-forward neural network autoregression) were used to predict COVID-19 metrics in the upcoming 14 days. The models were compared in terms of prediction ability using the root mean square error (RMSE) of the first principal component (PC1). The predictive abilities of models generated with both Google Trends data and conventional COVID-19 metrics were compared with those fitted with conventional COVID-19 metrics only.

Results: The degree of correlation and the best time lag varied as a function of the selected country and topic searched; in general, the optimal time lag was within 15 days. Overall, predictions of PC1 based on both search terms and COVID-19 traditional metrics performed better than those not including Google searches (median 1.56, IQR 0.90-2.49 versus median 1.87, IQR 1.09-2.95, respectively), but the improvement in prediction varied as a function of the selected country and time frame. The best model varied as a function of country, time range, and period of time selected. Models based on a 7-day moving average led to considerably smaller RMSE values as opposed to those calculated with raw data (median 0.90, IQR 0.50-1.53 versus median 2.27, IQR 1.62-3.74, respectively).

Conclusions: The inclusion of digital online searches in statistical models may improve the nowcasting and forecasting of the COVID-19 epidemic and could be used as one of the surveillance systems of COVID-19 disease. We provide a free web application operating with nearly real-time data that anyone can use to make predictions of outbreaks, improve estimates of the dynamics of ongoing epidemics, and predict future or rebound waves.

(J Med Internet Res 2021;23(8):e28876) doi:10.2196/28876

KEYWORDS

Google Trends; symptoms; coronavirus; SARS-CoV-2; big data; time series; predictive models; Shiny web application; infodemiology; infoveillance; digital health; COVID-19

Introduction

COVID-19 is a new entity, and the dynamics of its propagation are difficult to predict. In the absence of compelling evidence, health and political decisions have been strongly driven by a wide variety of statistical models and simulation scenarios to forecast the COVID-19 epidemic. Still, large variations exist among the different models with respect to the predicted number of infected people, time to reach a peak of new cases, course of the epidemic, and identification of outbreaks [1]. One key limitation of such models is that they rely heavily on the number of confirmed infected subjects who usually seek medical attention due to moderate to severe symptoms. However, confirmed cases are most likely only a small proportion of the true number of cases as the vast majority of infected individuals are asymptomatic or mildly symptomatic [2].

There is increasing interest in the potential of “big data” analysis to predict future areas of COVID-19 outbreaks and incidence of cases based on symptom search behaviors. In the past, search query data have been used to facilitate early detection and near real-time estimates of flu and Dengue [3]. A few studies have shown a correlation between Google Trends data of medical term searches and COVID-19 metrics [4], suggesting that incorporating Google Trends data into conventional metrics could lead to better nowcasting and forecasting of the COVID-19 epidemic.

In this study, we systematically evaluate patterns of web queries for COVID-19 clinical manifestations and develop an open-access web application for exploring their correlations with COVID-19 propagation. We implement models integrating conventional COVID-19 metrics with Google Trends data and compare them to those not containing Google Trends data. The aim of this study is to present a framework for digital surveillance of COVID-19 using open-access big data of Google searches of symptoms associated with COVID-19.

Methods

Data Collection

Daily new confirmed COVID-19 cases, the cumulative number of COVID-19 cases, and the number of cases and deaths per million for all available countries were exported from the COVID-19 Data Repository by the Center for System Science and Engineering at Johns Hopkins University [5]. The selected countries used as case studies are given in the results section below. Country choice was arbitrary, and the following principles were adopted: representation of the five continents; inclusion of countries where the COVID-19 epidemic had different levels of severity and different evolutions over time; inclusion of countries where Google is the preferred search engine; exclusion of countries with limited access to the internet; exclusion of countries where one or more Google Trends topic had only zero or missing values in the selected time frame; and exclusion of countries whose reliability in terms of data reporting has been questioned. As data were fully anonymized and publicly available, no ethical approval was required.

The Google Trends application programming interface (API) was used to extract trends of Google searches for the most common COVID-19 signs and symptoms. For each search term, geographic region, and time frame selected, Google Trends outputs an “interest-over-time” index, which estimates the relative search volume on a normalized scale from 0 (no searches) to 100 (search term popularity peak). A total of 20 topics were identified on the basis of the most frequent signs and symptoms of COVID-19 and included the following: abdominal pain, ageusia, anorexia, anosmia, bone pain, chills, conjunctivitis, cough, diarrhea, eye pain, fatigue, fever, headache, myalgia, nasal congestion, nausea, rhinorrhea, shortness of breath, sore throat, and tearing [6-9]. Google Trends queries were carried out with the “topic” function, which includes all the related terms sharing the same concept in different languages. This approach ensures that the frequency of searches for closely related symptom types are appropriately grouped together.

For each country and search term, data were automatically exported as CSV files for two prespecified time frames: (1) five years of weekly data from July 1, 2015, to July 1, 2020, to study the long-term pattern of search terms, and (2) daily data from January 22, 2020, to December 20, 2020. As Google Trends allows daily data exportation up to 9 months, daily data were reconstructed by means of an overlapping method [10].

Data Analysis

Interest-over-time values for the 5-year interval were used to distinguish topics with a significant deviation from their long-term pattern from the onset of the COVID-19 epidemic. For seasonal queries, trends were isolated from seasonal and random components with an additive decomposition method (Figure S1 in Multimedia Appendix 1); for nonseasonal queries, trends were extracted by smoothing the time series with a 1-year moving average. Decomposition plots were visually inspected, and topics with no clear change in their 5-year trends from January 2020 were excluded from the subsequent analyses.

The relationship between the daily interest-over-time values for the selected topics and COVID-19 confirmed deaths and new cases were investigated in the shorter time frame indicated above. Relationships between interest-over-time values of each topic and the number of new daily confirmed cases or deaths per million were visually assessed with line graphs. Changes in interest-over-time values over time were visually assessed with streamgraphs. To smooth daily fluctuations in both interest-over-time values and number of new cases, plots were generated using a 7-day moving average.

Time-lagged cross-correlations between COVID-19 new cases and each topic were calculated, using a 7-day moving average of both interest-over-time values and COVID-19 confirmed cases and deaths to blunt the day-by-day fluctuation.

Model Development and Assessment

Principal component analysis (PCA) was used to perform data dimensionality reduction, decrease the number of input variables, and filter out noisy or redundant information. For each country, two PCA models were implemented: one using unprocessed data and the other using 7-day moving average smoothing. PCA was applied to standardized data (ie, with zero mean and unit variance). The PCA model was graphically inspected through PCA score and loading plots. PCA was assessed via 5-fold cross-validation, and the results obtained in each test sample were averaged. The amount of variance explained by each principal component (PC) in the model was inspected with scree plots. Based on the elbow and Kaiser rules, the first two PCs (PC1 and PC2) were subsequently used for time series modelling [11].

A total of three different time series models were fitted on PC1 and PC2 values: error, trend, seasonality (ETS); autoregressive integrated moving average (ARIMA); and a feed-forward neural network autoregression (NNAR) model with one hidden layer [12]. Models were fitted on a 30-day window and used to predict future PC1 and PC2 values up to 14 days. The 14th predicted day was aligned to the peak and base of each wave. The new data scores predicted with the time series models were then reinserted into the model as input variables.

For each country, the three models were compared in terms of ability to predict the PC1 and PC2 using the root mean square error (RMSE) of the predicted values. For each time series model, the predictive abilities of the model generated with raw data and the one generated with 7-day moving averages were compared. To further assess the PCA models based on both Google Trends data and conventional COVID-19 metrics, we also generated predictive models based on conventional COVID-19 metrics only; we then compared the predictive ability of models with and without Google Trends data by means of RMSE for each country.

Web Application

An open-source web application was developed using R Shiny [13]. Data are collected, imported, and updated daily for 188 countries from the sources mentioned above.

The web application allows users to generate line graphs and streamgraphs to visualize interest-over-time values and COVID-19 metrics and view worldwide trends over time in a choropleth map. Relationships between the variables at various lags can be explored with cross-correlations. The web application allows fitting and evaluating PCA models, fitting a time series model (either ETS or ARIMA), predicting PC components or any of the input variables of the model (including numbers of new cases and deaths), and evaluating the model performance graphically and with various metrics, such as RMSE and mean absolute error. The user has operational control of several model features, including the subset of variables to build the PCA model, the time window to fit the time series model, and the time interval to predict.

Data Sharing

Raw data used in this study are publicly available on the COVID-19 Data Repository by the Center for System Science and Engineering at Johns Hopkins University [14] and the Google Trends webpage [15]. All processed data used in this study can be accessed and reproduced on the PredictPandemic website [13]. The code for the current version of the web application is open source and freely available [16].

Results

For this study, three European countries (Italy, United Kingdom, and France), one Asian country (India), one Oceanian country (Australia), one North American country (United States), one South American country (Brazil), one African country (South Africa), and one Middle Eastern country (Iran) were chosen as case studies. The cumulative numbers of cases and deaths in the selected countries are illustrated in Figure 1.

Table 1 summarizes information from the 5-year analysis. Among the 20 screened topics, 13 showed seasonality, while the remaining were nonseasonal. Overall, 11 search topics (Figures S2-S20 in Multimedia Appendix 1) showed a clear deviation from their 5-year trend: ageusia, anosmia, chills, cough, eye pain, fever, headache, nasal congestion, rhinorrhea, shortness of breath, and sore throat.

Figure 1. Cumulative number of confirmed cases (A) and deaths (B) per million for each country over time.

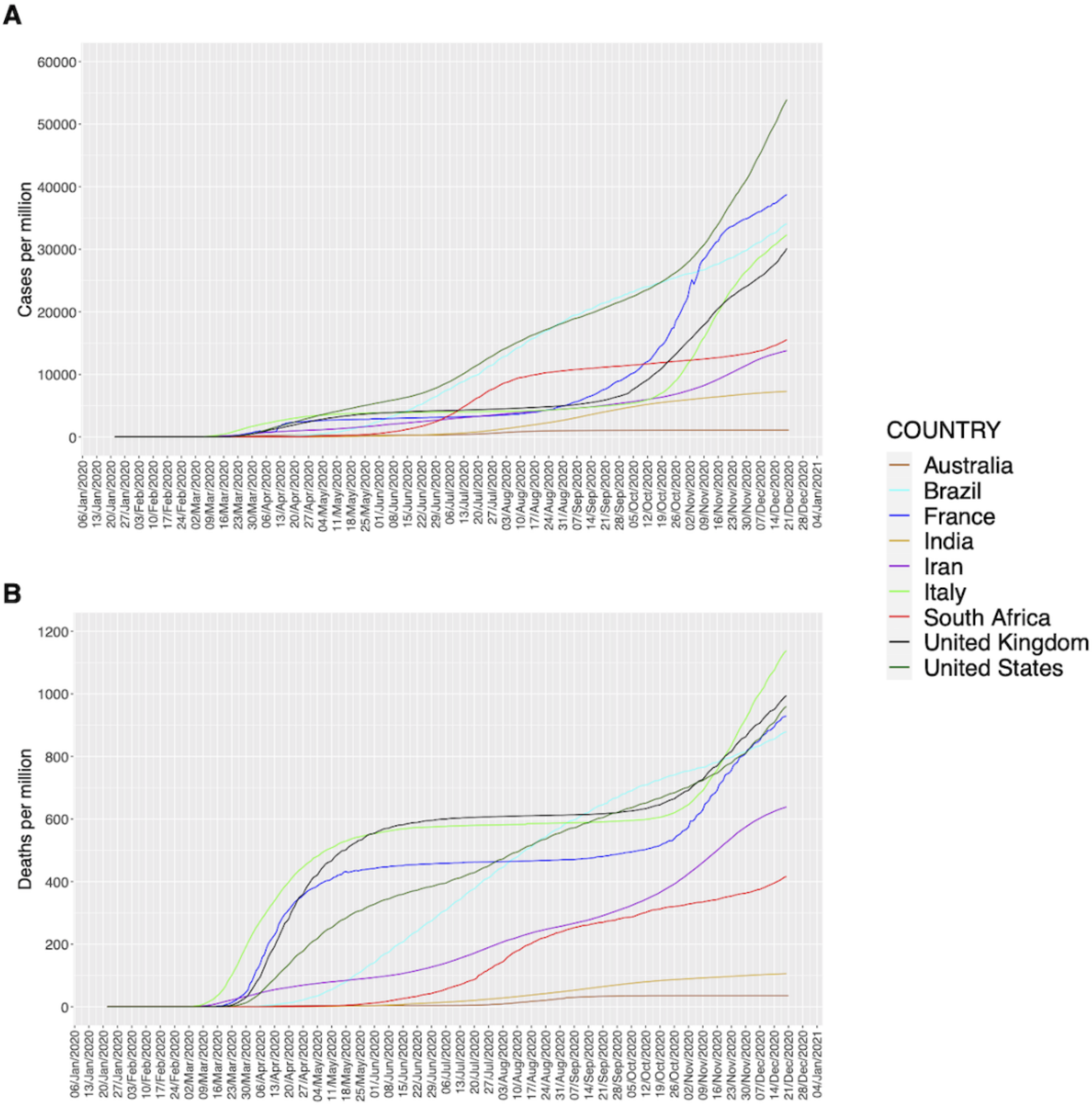


Table 1. Symptoms screened in the 5-year analysis.

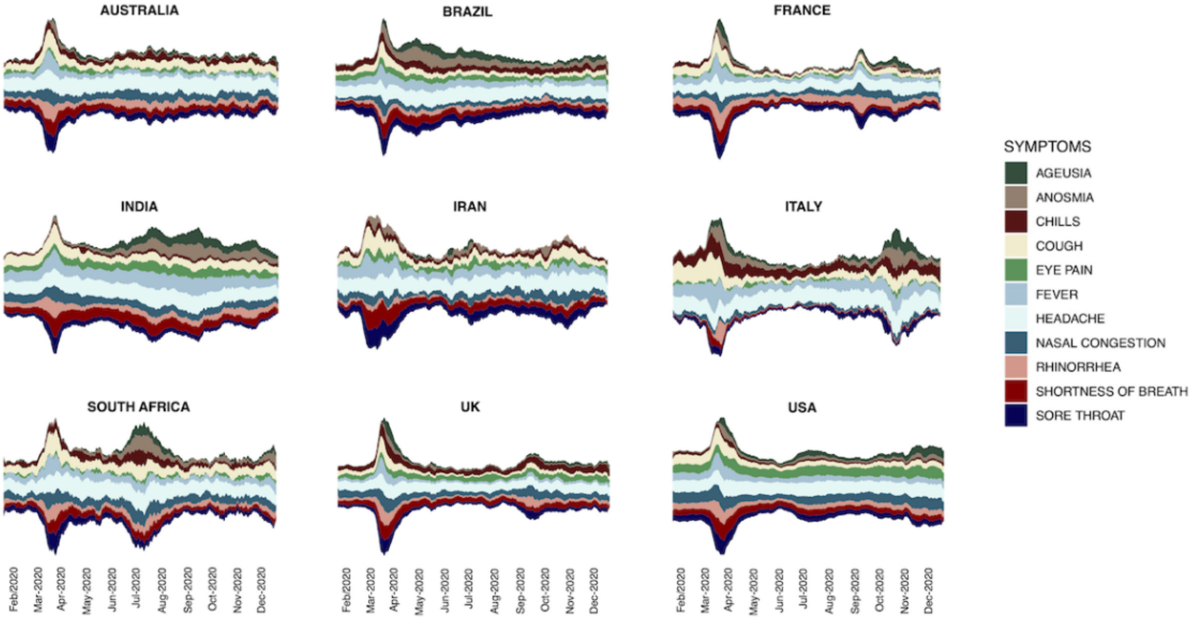
Topic	Seasonality	Deviation from 5-year trend[a]
Abdominal pain	Nonseasonal	No
Ageusia	Nonseasonal	Yes
Anorexia	Seasonal	No
Anosmia	Nonseasonal	Yes
Bone pain	Nonseasonal	No
Chills	Seasonal	Yes
Conjunctivitis	Seasonal	No
Cough	Seasonal	Yes
Diarrhea	Seasonal	No
Eye pain	Nonseasonal	Yes
Fatigue	Seasonal	No
Fever	Seasonal	Yes
Headache	Seasonal	Yes
Myalgia	Seasonal	No
Nasal congestion	Seasonal	Yes
Nausea	Nonseasonal	No
Rhinorrhea	Seasonal	Yes
Shortness of breath	Seasonal	Yes
Sore throat	Seasonal	Yes
Tearing	Nonseasonal	No

[a] Topic categorization into deviating and not deviating from their 5-year trend was determined on the visual inspection of decomposition plots.

The relationships between the number of new cases and each search topic are illustrated in Figures S21-S29 in Multimedia Appendix 1. Several symptoms, including ageusia, anosmia, cough, rhinorrhea, and sore throat were aligned with the COVID-19 epidemic in most countries and were searched on Google well before the number of COVID-19 confirmed cases peaked. On the other hand, other topics showed less evident variations (chills, eye pain) or deviated from their trend only during the first wave (headache, shortness of breath). Additionally, the peak of interest in all symptoms (except eye pain) preceded that of confirmed COVID-19 cases in most countries, and topics increasing earlier reached their highest interest-over-time value before those growing later. Similar patterns were observed for interest-over-time of search terms when compared to the number of newly confirmed deaths (Figures S30-S38 in Multimedia Appendix 1).

The interest-over-time change for all topics is illustrated in Figure 2. Overall, the interest-over-time values of the selected topics had a peak in March 2020 in all selected countries. In Italy, France, and South Africa—and, to a lesser extent, Iran, the United Kingdom, and the United States—there was a decrease in the search terms after the first peak, followed by a second peak. In Iran, a third peak in searches was seen, corresponding to the third COVID-19 wave. In India and Brazil, searches of medical terms remained high after the first peak, and no second peak was seen. In Australia, the interest-over-time values of the selected topics returned to the pre-peak values soon after the first peak in March and remained low and stable.

Figure 2. Streamgraphs of the interest-over-time index for each individual country. The x-axis values are given in months. Index-over-time values were plotted as 7-day moving average.

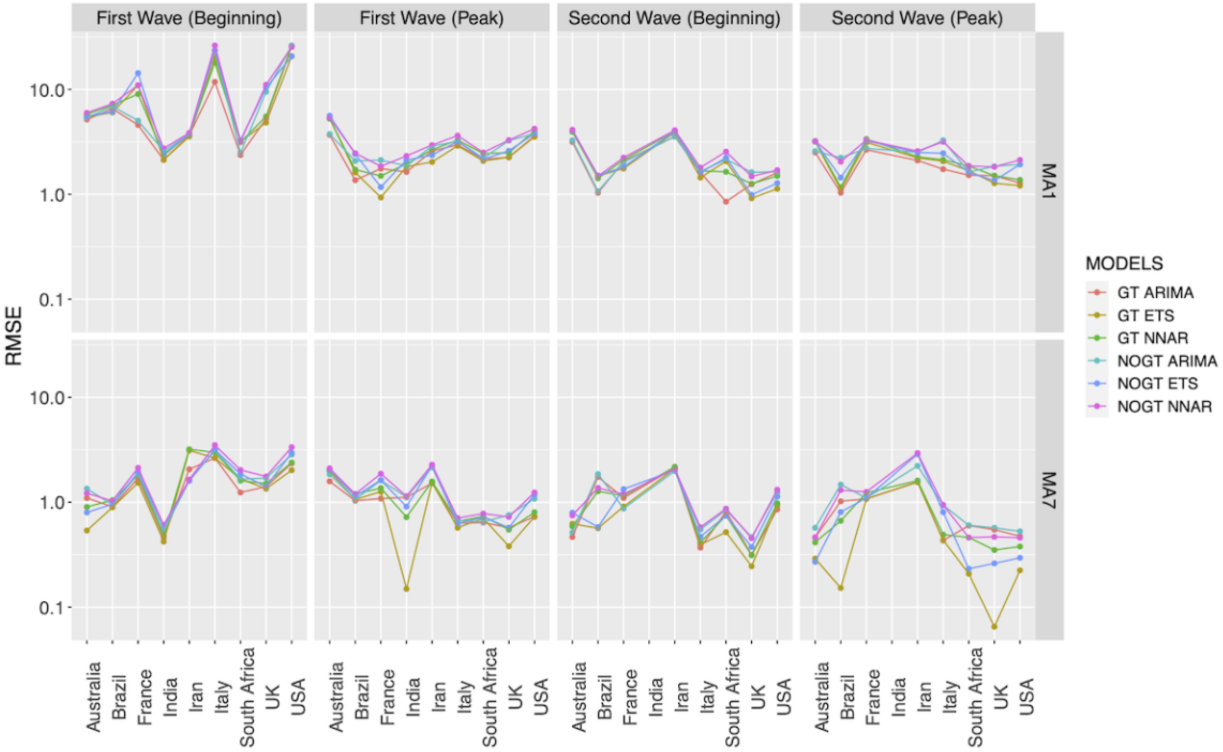


Cross-correlations between each topic and the number of confirmed COVID-19 cases are reported in Tables S1-S9 in Multimedia Appendix 1. Overall, ageusia, anosmia, and headache were most consistently correlated with COVID-19 cases across the selected countries. The degree of correlation and the best time lag largely varied as a function of the selected country and topic, but overall the optimal time lag was 15 days.

The scores and loadings plots for the PCA models are given in Figures S39-S47 in Multimedia Appendix 1. The scores plot represents a summary of the collected data trends over time, while the loadings plot shows how strongly each variable influences a PC. In the month of March 2020, all selected countries deviated considerably from their previous scores and moved toward the PCs directions of the loadings of the Google search terms, preceding the increment in the number of searches of the symptoms related to COVID-19. The latest 14 days show a similar pattern for all selected countries, pointing toward the loadings' directions of deaths and new cases. Specifically, for Italy, France, and South Africa, the latest scores are in the same area of loadings plots corresponding to deaths and new cases, indicating a stable trend in these metrics. On the other hand, the United Kingdom, the United States, Brazil, and India followed a worsening pattern, as their scores kept moving toward the direction of new deaths and cases. Australia and Iran were the only selected countries showing an improving trend, with the score points moving away from the loadings of COVID-19 deaths and new cases.

Figure 3 illustrates the RMSE for the prediction of PC1 values with the three time series models in the various countries using raw data and a 7-day moving average. Models based on the 7-day moving average lead to considerably smaller RMSE values as opposed to those calculated with raw data (median 0.90, IQR 0.50-1.53 versus median 2.27, IQR 1.62-3.74, respectively). Overall, predictions based on both search terms and COVID-19 conventional metrics performed better than those not including Google searches (median 1.56, IQR 0.90-2.49 versus median 1.87, IQR 1.09-2.95, respectively), but the improvement in prediction varied as a function of the selected country and time frame. Although ETS (median 1.62, IQR 0.87-2.7) led to slightly smaller RMSE than ARIMA (median 1.65, IQR 1.04-2.58) and NNAR (median 1.82, IQR 1.15-3.15) models, none of the tested time series models clearly outperformed the remaining two, and the best model varied as a function of country, time range, and period of time selected. Similar results were obtained when trying to predict PC2 (Figure S48 in Multimedia Appendix 1).

Figure 3. Root mean square errors of the prediction error for principal component 1 of the various models for the selected countries. MA1 and MA7 indicate analyses performed on 1-day (ie, original data) and 7-day moving averages of data, respectively. GT indicates models based on both traditional COVID-19 metrics and Google Trends data, while NOGT models are based on COVID-19 metrics only. ARIMA: autoregressive integrated moving average; ETS: error, trend, seasonality; NNAR: feed-forward neural network autoregression; RMSE: root mean square error; UK: United Kingdom; USA: United States of America.



Discussion

In this study, we investigated the relationship between Google Trends searches of symptoms associated with COVID-19 and confirmed COVID-19 cases and deaths. We found that some of the search terms showed an unusually high recent online interest that deviated considerably from their expected behavior and preceded the peak of confirmed COVID-19 cases by days to weeks. This pattern was consistent across different countries and of similar magnitude. We developed and validated predictive models to forecast the COVID-19 epidemic based on the combination of Google Trends searches of symptoms associated with COVID-19 and traditional COVID-19 metrics. We found that models incorporating Google Trends data generally performed better than those based solely on traditional COVID-19 metrics. We also developed a web application [13] to translate our approach into action.

Our study identified patterns of Google searches of several symptoms and signs associated with COVID-19 in a consistent way across the studied countries. Overall, Google searches of COVID-19 symptoms followed a similar trend to that of the COVID-19 epidemic and preceded traditional COVID-19 metrics. This behavior can contribute to the early recognition of new waves and epidemic peaks.

The interpretation of symptom search behavior during COVID-19 outbreaks should be carefully considered. Dynamics of online searches may show atypical patterns during pandemics where major restrictions occur, including shutdowns of economic activities, movement restrictions, and health care overload [17]. Constant media attention may contribute to increasing interest in some of the studied topics [18]. COVID-19 received extensive coverage that might have precipitated unusually high interest during lockdowns [19]. Our findings on online search behavior might be secondary to general media interest in specific COVID-19 symptoms, rather than a primary, and possibly predictive, consequence of people with COVID-19 researching their own symptoms in real time. All the selected countries had a peak in searches of medical terms in or around March and April 2020, including those countries with low numbers of cases at that time, such as South Africa and India. This pattern may indicate that curiosity and media coverage regarding the new pandemic can explain part of the first peak in search terms, in agreement with a previous study [20]. After the first peak, however, Google search behavior followed different patterns across countries and resembled the course of the COVID-19 epidemic. In those countries that had a second wave—such as Italy, France, Iran, the United Kingdom, or the United States—the number of Google searches had a second peak; the height of the second peak of the searches was lower than that of the first peak, despite a higher number of reported cases and deaths, suggesting that individual curiosity regarding the new pandemic could have inflated the first peak in search terms. Iran also had a third peak in searches between October and December 2020, when the country had a third COVID-19 wave. South Africa had a second peak in its searches in July 2020, when the country had its first wave. In India, the first peak in searches was followed by a steady increase from June 2020, after which searches remained stable until October 2020, and then decreased gradually, resembling the shape of the

COVID-19 epidemic in this country. In Australia, which effectively managed the COVID-19 epidemic and had among the lowest infection and death rates in the world, the interest-over-time of the various search terms after the first peak remained low and comparable to pre-peak values.

We observe that not all the selected topics reached their peak searches simultaneously; rather, they had different time patterns, which were fairly consistent across all countries. We believe that the intense and simultaneous media coverage of all the selected topics should have had the same effect at the same date if the media influence entirely caused this search behavior [21]. Ageusia and anosmia showed the highest correlations when lagged by a few days, while cough, fever, nasal congestion, sore throat, rhinorrhea, and shortness of breath preceded increases in COVID-19 cases by up to two weeks. This finding is consistent with the clinical course of COVID-19; in a large multicenter European study, olfactory and gustatory dysfunctions were among the latest and first manifestations in approximately 65% and 12% of patients, respectively [8].

In addition to describing how Google search terms changed over time in different countries and investigating their relationship with the numbers of cases and deaths, we also developed models combining interest-over-time values of searches of COVID-19 symptoms with conventional metrics (eg, number of new cases, number of new deaths) to predict the course of the COVID-19 epidemic, and we compared the prediction ability of these models against that of models based only on conventional metrics. The PCA approach allowed us to reduce dimensionality, summarize information into 2 PCs, and filter out the noisy or redundant information. Another advantage of PCA was to provide visual representations of data patterns, similarity trends, and outliers. The PCA approach is highly flexible and potentially allows accommodating new variables of interest in future versions of our application. As the PCA itself does not make any predictions, we processed the PC computed values with different time series models, and new data scores predicted with the time series models were reinserted into the PCA model. Our approach allows the extraction of the predicted values of any input variable, including the number of new cases and deaths. Models integrating interest-over-time values of the search topics and COVID-19 traditional metrics generally outperformed models based solely on confirmed cases and deaths, leading to improved predictions. There was no single best model in this study, and the best performing time series model varied as a function of the country, time frame, and moving average. Predictions were more accurate, leading to considerably smaller RMSE, when obtained using a 7-day moving average rather than daily data. This result is not surprising as Google Trends data have high daily fluctuations, and COVID-19 reported cases greatly oscillate, reflecting testing and reporting practices and contingencies [22,23].

To translate our results into practice so that the scientific community, agencies, and even curious users could potentially use them, we developed a freely available web application [13]. The application is interactive and updates the data daily, so it operates in near real time. It allows the user to visualize data for 188 countries, choosing any time frame. In addition, COVID-19 traditional metrics and Google search terms' interest-over-time can be visualized globally on different graphs. The user can explore cross-correlations among selected keywords, generate predictive models with default variables or a user-selected subset of variables, and check model performance.

This study has limitations. The Google Trends algorithm is a "black box," and the exact calculation formula for interest over time and raw data has never been made public. Search results may differ slightly when downloaded by different computers or on different days. However, we conducted search-research reliability, which showed excellent reliability for most of the topics included in this study (data not shown). The exclusion of those symptoms with no significant deviation from their 5-year trend reduced the possibility of spurious correlations, but it was not possible to account for seasonality in the selected topics. In other words, a small proportion of the increasing trend in some topics might be explained by their usual seasonal variations. The results of this study may not apply to countries where Google is not a popular search engine or where Google is censored or limited in its use. However, this approach can be applied to other search engines (eg, Baidu, Yahoo, Naver), as was done in previous studies on different diseases and on COVID-19 in Hubei province, China [24,25]. Previous studies have shown that self-reported symptoms on social media networks, such as Twitter, can provide useful information to track the COVID-19 pandemic and can be used for infoveillance along with search engine data [26-28]. Our approach could, in principle, be applied to social media as well. Geographical areas and groups of people (older adults and children) with scarce internet access cannot be studied with this strategy, and our results may not apply to largely rural countries. This study included only the most common clinical manifestations of COVID-19, and only a few selected countries were included as a case study. However, information and models for every country can be found on our web application.

Future work will include increased data granularity, allowing users to access information and make predictions at a regional level. Other metrics of interest, such as hospitalizations, will be included in our analysis as outcomes. Finally, we plan to allow the user to generate a one-page report for each country, summarizing the most relevant information.

In conclusion, the results of this study show that Google Trends searches during the COVID-19 pandemic may precede outbreaks by up to two weeks. The inclusion of digital online searches in statistical models may improve the nowcasting and forecasting of the COVID-19 epidemic and could be used as one of the surveillance systems employed by government agencies and supranational organizations to refine their monitoring of COVID-19. We provide a free web application operating with nearly real-time data that anyone can use to make predictions of outbreaks, improve estimates of dynamics of ongoing epidemics, and predict future or rebound waves.

Acknowledgments

This study was supported by the EOSCsecretariat.eu, which has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-05-2018-2019, grant Agreement number 831644.

Authors' Contributions

AR and AM contributed to study conception, design, data acquisition, data analysis, data interpretation, manuscript drafting, and manuscript revision. EA contributed to data analysis, data interpretation, manuscript drafting, and manuscript revision. EM contributed to data acquisition, data analysis, and manuscript revision. AAA, AIM, and FB contributed to data interpretation and manuscript revision. All authors had full access to all the data in the study. All authors revised the manuscript and approved the final version before submission. The corresponding author had final responsibility for the decision to submit for publication.

Conflicts of Interest

AIM reports personal fees and consultancy fees from Allergan, Pfizer, Alcon, Novartis, Zeiss, Easyscan, and Visufarma, outside the submitted work. FB reports consultancy fees from Allergan, Bayer, Boehringer-Ingelheim, Fidia Sooft, Hofmann La Roche, Novartis, NTC Pharma, Sifi, Thrombogenics, and Zeiss. All other authors declare no competing interests.

Multimedia Appendix 1
Supplementary figures and tables.
[DOCX File , 8955 KB - jmir_v23i8e28876_app1.docx]

References

1. Roda WC, Varughese MB, Han D, Li MY. Why is it difficult to accurately predict the COVID-19 epidemic? Infect Dis Model 2020;5:271-281 [FREE Full text] [doi: 10.1016/j.idm.2020.03.001] [Medline: 32289100]
2. Fu L, Wang B, Yuan T, Chen X, Ao Y, Fitzpatrick T, et al. Clinical characteristics of coronavirus disease 2019 (COVID-19) in China: A systematic review and meta-analysis. J Infect 2020 Jun;80(6):656-665 [FREE Full text] [doi: 10.1016/j.jinf.2020.03.041] [Medline: 32283155]
3. 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]
4. Walker A, Hopkins C, Surda P. Use of Google Trends to investigate loss-of-smell-related searches during the COVID-19 outbreak. Int Forum Allergy Rhinol 2020 Jul;10(7):839-847 [FREE Full text] [doi: 10.1002/alr.22580] [Medline: 32279437]
5. 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 [FREE Full text] [doi: 10.1016/S1473-3099(20)30120-1] [Medline: 32087114]
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. Fu L, Wang B, Yuan T, Chen X, Ao Y, Fitzpatrick T, et al. Clinical characteristics of coronavirus disease 2019 (COVID-19) in China: A systematic review and meta-analysis. J Infect 2020 Jun;80(6):656-665 [FREE Full text] [doi: 10.1016/j.jinf.2020.03.041] [Medline: 32283155]
8. Lechien JR, Chiesa-Estomba CM, De Siati DR, Horoi M, Le Bon SD, Rodriguez A, et al. Olfactory and gustatory dysfunctions as a clinical presentation of mild-to-moderate forms of the coronavirus disease (COVID-19): a multicenter European study. Eur Arch Otorhinolaryngol 2020 Aug;277(8):2251-2261 [FREE Full text] [doi: 10.1007/s00405-020-05965-1] [Medline: 32253535]
9. Wu P, Duan F, Luo C, Liu Q, Qu X, Liang L, et al. Characteristics of Ocular Findings of Patients With Coronavirus Disease 2019 (COVID-19) in Hubei Province, China. JAMA Ophthalmol 2020 May 01;138(5):575-578 [FREE Full text] [doi: 10.1001/jamaophthalmol.2020.1291] [Medline: 32232433]
10. Tseng Q. Reconstruct Google Trends Daily Data for Extended Period. towards data science. 2019. URL: https://towardsdatascience.com/reconstruct-google-trends-daily-data-for-extended-period-75b6ca1d3420 [accessed 2021-07-16]
11. Afifi A, May S, Donatello R, Clark V. Practical Multivariate Analysis (6th ed). Boca Raton, FL: Chapman & Hall/CRC; 2020.
12. Hyndman RJ, Khandakar Y. Automatic Time Series Forecasting: The forecast Package for R. J Stat Soft 2008;27(3):v027i03. [doi: 10.18637/jss.v027.i03]
13. Rabiolo A, Alladio E, Morales E, Marchese A. PredictPandemic.org. URL: https://predictpandemic.org [accessed 2021-07-16]
14. Center for System Science and Engineering at Johns Hopkins University. COVID-19 Data Repository. URL: https://github.com/CSSEGISandData/COVID-19 [accessed 2021-07-16]
15. Google Trends. URL: https://trends.google.com [accessed 2021-07-15]
16. Rabiolo A, Alladio E, Morales E, Marchese A. PredictPandemic web-application. URL: https://zenodo.org/record/4603713#.YE4sfC2l28U [accessed 2021-07-16]

17. Adebayo G, Neumark Y, Gesser-Edelsburg A, Abu Ahmad W, Levine H. Zika pandemic online trends, incidence and health risk communication: a time trend study. BMJ Glob Health 2017;2(3):e000296 [FREE Full text] [doi: 10.1136/bmjgh-2017-000296] [Medline: 29082006]
18. Cervellin G, Comelli I, Lippi G. Is Google Trends a reliable tool for digital epidemiology? Insights from different clinical settings. J Epidemiol Glob Health 2017 Sep;7(3):185-189 [FREE Full text] [doi: 10.1016/j.jegh.2017.06.001] [Medline: 28756828]
19. Effenberger M, Kronbichler A, Shin JI, Mayer G, Tilg H, Perco P. Association of the COVID-19 pandemic with Internet Search Volumes: A Google Trends Analysis. Int J Infect Dis 2020 Jun;95:192-197 [FREE Full text] [doi: 10.1016/j.ijid.2020.04.033] [Medline: 32305520]
20. Szmuda T, Ali S, Hetzger TV, Rosvall P, Słoniewski P. Are online searches for the novel coronavirus (COVID-19) related to media or epidemiology? A cross-sectional study. Int J Infect Dis 2020 Aug;97:386-390 [FREE Full text] [doi: 10.1016/j.ijid.2020.06.028] [Medline: 32535297]
21. Adawi M, Bragazzi NL, Watad A, Sharif K, Amital H, Mahroum N. Discrepancies Between Classic and Digital Epidemiology in Searching for the Mayaro Virus: Preliminary Qualitative and Quantitative Analysis of Google Trends. JMIR Public Health Surveill 2017 Dec 01;3(4):e93 [FREE Full text] [doi: 10.2196/publichealth.9136] [Medline: 29196278]
22. Bergman A, Sella Y, Agre P, Casadevall A. Oscillations in U.S. COVID-19 Incidence and Mortality Data Reflect Diagnostic and Reporting Factors. mSystems 2020 Jul 14;5(4):00544-20 [FREE Full text] [doi: 10.1128/mSystems.00544-20] [Medline: 32665331]
23. Rovetta A. Reliability of Google Trends: Analysis of the Limits and Potential of Web Infoveillance During COVID-19 Pandemic and for Future Research. Front Res Metr Anal 2021;6:670226 [FREE Full text] [doi: 10.3389/frma.2021.670226] [Medline: 34113751]
24. 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]
25. Qiu HJ, Yuan LX, Huang XK, Zhou YQ, Wu QW, Zheng R, et al. [Using the big data of internet to understand the characteristics of coronavirus disease 2019: a big data study]. Zhonghua Er Bi Yan Hou Tou Jing Wai Ke Za Zhi 2020 Jun 07;55(6):569-575. [doi: 10.3760/cma.j.cn115330-20200225-00128] [Medline: 32186171]
26. Klein AZ, Magge A, O'Connor K, Flores Amaro JI, Weissenbacher D, Gonzalez Hernandez G. Toward Using Twitter for Tracking COVID-19: A Natural Language Processing Pipeline and Exploratory Data Set. J Med Internet Res 2021 Jan 22;23(1):e25314 [FREE Full text] [doi: 10.2196/25314] [Medline: 33449904]
27. Mackey T, Purushothaman V, Li J, Shah N, Nali M, Bardier C, et al. Machine Learning to Detect Self-Reporting of Symptoms, Testing Access, and Recovery Associated With COVID-19 on Twitter: Retrospective Big Data Infoveillance Study. JMIR Public Health Surveill 2020 Jun 08;6(2):e19509 [FREE Full text] [doi: 10.2196/19509] [Medline: 32490846]
28. Panuganti BA, Jafari A, MacDonald B, DeConde AS. Predicting COVID-19 Incidence Using Anosmia and Other COVID-19 Symptomatology: Preliminary Analysis Using Google and Twitter. Otolaryngol Head Neck Surg 2020 Sep 02;163(3):491-497 [FREE Full text] [doi: 10.1177/0194599820932128] [Medline: 32484425]

Abbreviations

ARIMA: autoregressive integrated moving average
ETS: error, trend, seasonality
NNAR: neural network autoregression
PC: principal component
PCA: principal component analysis

Edited by C Basch; submitted 17.03.21; peer-reviewed by P Perco, B Wang, M Effenberger; comments to author 08.04.21; revised version received 28.04.21; accepted 01.05.21; published 11.08.21.

Please cite as:
Rabiolo A, Alladio E, Morales E, McNaught AI, Bandello F, Afifi AA, Marchese A
Forecasting the COVID-19 Epidemic by Integrating Symptom Search Behavior Into Predictive Models: Infoveillance Study
J Med Internet Res 2021;23(8):e28876
URL: https://www.jmir.org/2021/8/e28876
doi:10.2196/28876
PMID:34156966

©Alessandro Rabiolo, Eugenio Alladio, Esteban Morales, Andrew Ian McNaught, Francesco Bandello, Abdelmonem A Afifi, Alessandro Marchese. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 11.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Using Infodemiology Metrics to Assess Public Interest in Liver Transplantation: Google Trends Analysis

Maria Effenberger[1], MD; Andreas Kronbichler[2], PhD; Erica Bettac[3], BSc; Felix Grabherr[1], MD; Christoph Grander[1], MD; Timon Erik Adolph[1], PhD; Gert Mayer[2], MD; Heinz Zoller[1], MD; Paul Perco[2], PhD; Herbert Tilg[1], MD

[1]Department of Internal Medicine I, Gastroenterology, Hepatology, Endocrinology and Metabolism, Medical University of Innsbruck, Innsbruck, Austria
[2]Department of Internal Medicine IV, Nephrology and Hypertensiology, Medical University of Innsbruck, Innsbruck, Austria
[3]Department of Psychology, Washington State University Vancouver, Vancouver, WA, United States

Corresponding Author:
Herbert Tilg, MD
Department of Internal Medicine I, Gastroenterology, Hepatology, Endocrinology and Metabolism
Medical University of Innsbruck
Anichstrasse 35
Innsbruck, 6020
Austria
Phone: 43 512 504 23539
Fax: 43 512 504 23538
Email: herbert.tilg@i-med.ac.at

Abstract

Background: Liver transplantation (LT) is the only curative treatment for end-stage liver disease. Less than 10% of global transplantation needs are met worldwide, and the need for LT is still increasing. The death rates on the waiting list remain too high.

Objective: It is, therefore, critical to raise awareness among the public and health care providers and in turn increasingly acquire donors.

Methods: We performed a Google Trends search using the search terms *liver transplantation* and *liver transplant* on October 15, 2020. On the basis of the resulting monthly data, the annual average Google Trends indices were calculated for the years 2004 to 2018. We not only investigated the trend worldwide but also used data from the United Network for Organ Sharing (UNOS), Spain, and Eurotransplant. Using pairwise Spearman correlations, Google Trends indices were examined over time and compared with the total number of liver transplants retrieved from the respective official websites of UNOS, the Organización Nacional de Trasplantes, and Eurotransplant.

Results: From 2004 to 2018, there was a significant decrease in the worldwide Google Trends index from 78.2 in 2004 to 20.5 in 2018 (–71.2%). This trend was more evident in UNOS than in the Eurotransplant group. In the same period, the number of transplanted livers increased worldwide. The waiting list mortality rate was 31% for Eurotransplant and 29% for UNOS. However, in Spain, where there are excellent awareness programs, the Google Trends index remained stable over the years with comparable, increasing LT numbers but a significantly lower waiting list mortality (15%).

Conclusions: Public awareness in LT has decreased significantly over the past two decades. Therefore, novel awareness programs should be initialized.

(J Med Internet Res 2021;23(8):e21656) doi:10.2196/21656

KEYWORDS
digital medicine; search trends; public awareness; infodemiology; eHealth

Introduction

Background

Liver transplantation (LT) remains to be the only curative therapy for patients affected by end-stage liver disease, cirrhosis with hepatocellular carcinoma, acute fulminant hepatic failure, hepatocellular carcinoma, hilar cholangiocarcinoma, and several metabolic disorders [1,2].

Each year, approximately 12,000 LTs are performed in Europe and the United States, with numbers significantly increasing over time [3]. At present, more than 70% of liver transplant recipients survive for at least 5 years, compared with 20% in the 1980s. Such statistics are especially encouraging, considering that transplanted patients tend to have more severe diseases [4]. Several factors increase the survival of patients with LT, including better control of disease before LT, improved surgical techniques and surgeons specialized in these techniques, improved organ preservation, and advanced immunosuppressive therapy regimens [5]. However, the improved success rate of LT has resulted in substantial organ shortages [4]. Such shortages have led to a prolonged time for patients on the waiting list and increased waiting list mortality [6-8].

In 2017, the median pretransplant waiting time among active waitlisted adults was 10 months in the Eurotransplant region and approximately 9 months in the United States [9]. Mortality on the list was 18.7% in the Eurotransplant region, which is comparable with the United States' 19.8% of listed patient deaths before transplantation. It is important to note that limited donor organs are available from deceased donors after brain death [6].

The discrepancy between available liver allografts and transplant candidates continues to increase globally. Significant efforts have been made to raise the donor pool in Europe and the United States, such as using extended criteria donor organs [10], inventing extracorporeal normothermic or hypothermic organ perfusion systems [11], and accepting liver allografts as donation after circulatory determination of death (DCDD) [12]. Despite these efforts, there remains no significant decrease in waiting list mortality. To close the gap between available organs and the number of patients in need of LT, a higher awareness and acceptance of the transplant and donor program in the general population, as well as among health care providers, is a potentially effective strategy.

Infodemiology is an emerging area of research among health informatics, health care professionals, and patients. Introduced in 2002 [13], the term infodemiology is defined as a new area of scientific research that holds great promise for improving public health by focusing on specific internet searches for user-contributed health-related content [13-15]. These searches track public opinion, behavior, attention, knowledge, and attitudes [16].

Objectives

The first study indicated a correlation between searches on the internet and incidence in the field of *infectious diseases* [17]. The number of infodemiological studies has increased over the past decade, and these studies have used Twitter and Google [18]. Many researchers have used the infodemiological approach to study various health-related topics, for example, infectious diseases such as influenza or HIV/AIDS, chronic diseases such as multiple sclerosis, or patterns of smoking and tobacco use [19-27]. A particular interest in infodemiology has risen because of the ongoing COVID-19 pandemic [28], as research informs about the speed of misinformation [29], correlation between search behavior and COVID-19 related mortality [30], mental health issues [31], and pressing health care topics such as telehealth capacity of hospitals [32]. In addition, these data are becoming valuable tools for exploring human behavior. The advantage of infodemiology is that metrics are available in real time, which can provide quantitative and qualitative data while being automatically and inexpensively collected.

The analysis of internet search queries offers information on the extent of public attention, thereby reflecting the level of public awareness [33-36]. Google Trends is one of the most widely used tools for this purpose. It is not only used to study public interest in health care topics but also to predict disease occurrence and outbreaks [17,37,38].

In this study, we evaluated public interest in LT over time using Google Trends data and compared them with the number of transplanted livers reported from the United Network for Organ Sharing (UNOS), the Organización Nacional de Trasplantes (ONT), and the Eurotransplant regions.

Methods

Retrieving Transplantation Numbers for UNOS, ONT, and Eurotransplant

Data were retrieved by accessing the respective websites of the transplant organizations UNOS, ONT, and Eurotransplant [39-41]. We extracted information on living and deceased donors over a period of 15 years (2004-2018) for the following countries: the United States (UNOS), Spain (ONT), Austria, Belgium, Croatia, Germany, Hungary, Luxembourg, and the Netherlands (belonging to the Eurotransplant countries). No organs from executed prisoners were used in these transplant organizations.

Retrieving Google Trends Data on LT

The Google Trends tool was used on October 15, 2020, to retrieve data on internet user search activities in the context of LT [42]. Worldwide Google Trends indices were retrieved from January 2004 onward using the search terms, *liver transplantation* and *liver transplant*. We retrieved Google Trends indices for the United States, Spain, and European countries, in part included in the Eurotransplant network, namely Austria, Belgium, Croatia, Germany, Hungary, and the Netherlands. No Google Trends indices could be retrieved for Luxembourg and Slovenia. Whereas the worldwide search was performed in English, individual searches across non–English-speaking countries were performed in their respective official languages. We used individual search terms and combined the search terms yielding Google Trends results in larger queries, as listed in Table 1. On the basis of monthly data, annual average Google Trends indices were calculated for the years 2004 to 2018 and used to generate the line plots with the ggplot2 package of the statistical software R (version 3.4.1; R Foundation for Statistical Computing). It is important to note that none of the queries in the Google database for this study can be associated with a particular individual. The database does not retain information about the identity, IP address, or specific physical location of any user. The Spearman correlation coefficient was used to determine pairwise correlations between total liver transplant numbers per country and Google Trends indices.

Table 1. Google Trends search query listing (2004-present).

Region	Language	Google Trends search query
Worldwide	English	*liver transplantation* and *liver transplant*
United States	English	*liver transplantation* and *liver transplant*
Spain	Spanish	*trasplante de hígado*, *trasplante higado*, and *trasplante de higado*
Belgium	French	*transplantation hépatique* and *levertransplantatie*
The Netherlands	Dutch	*Levertransplantatie*
Germany	German	*Lebertransplantation*
Austria	German	*Lebertransplantation*
Hungary	Hungarian	*májátültetés* and *májtranszplantáció*
Croatia	Croatian	*transplantacija jetre*

Results

Google Trends and Trends for LT Worldwide

The global Google Trends index for LT decreased from 73.8 to 36.6 (–50.4%) between 2004 and 2014. In 2018, there was a slight upward trend in the LT index to 46.3 (+27.5%; Figure 1).

A similar trend was observed for the UNOS, with the Google Trends index dropping from 59.2 to 38.8 (–34.5%) in 2014 and an upward trend since then to 50.3 (+46.2%) until 2018. Similarly, Google Trends indices in the Eurotransplant region exhibited a decline in numbers in all Eurotransplant countries across the same period (Figure 2; Tables 2 and 3).

Figure 1. Worldwide interest in liver transplantation using Google Trends.

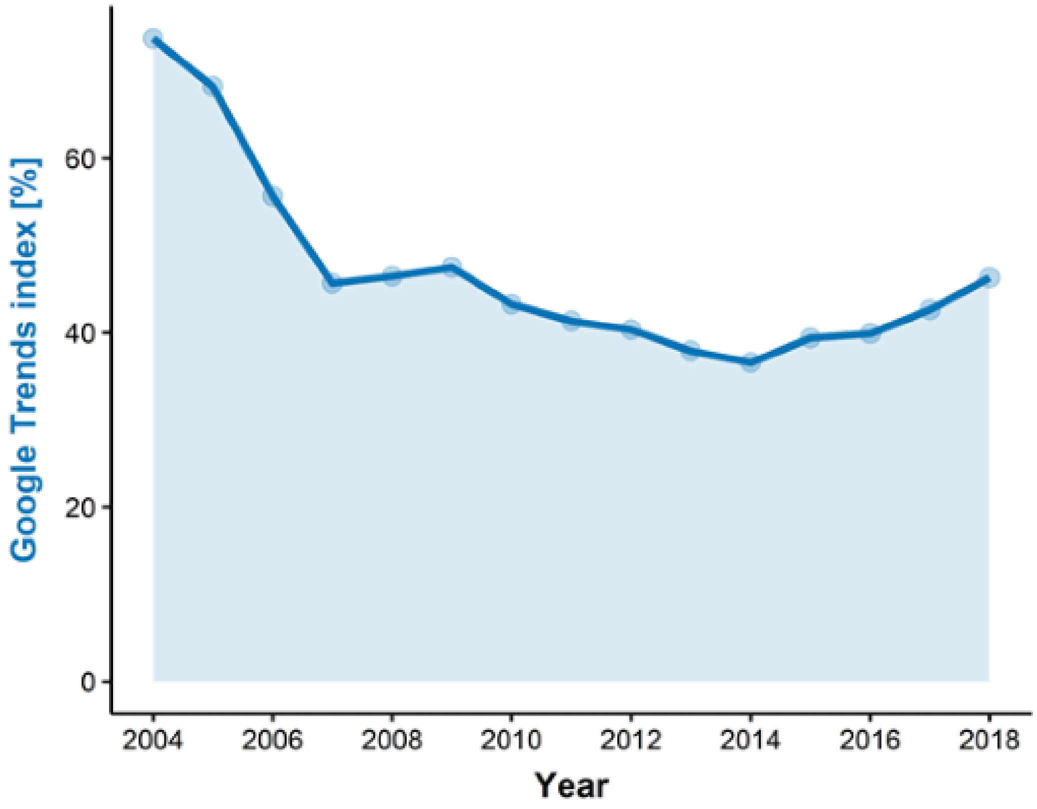


Figure 2. Google Trends and number of liver transplants in Eurotransplant, United Network for Organ Sharing, and Organización Nacional de Trasplantes over time.

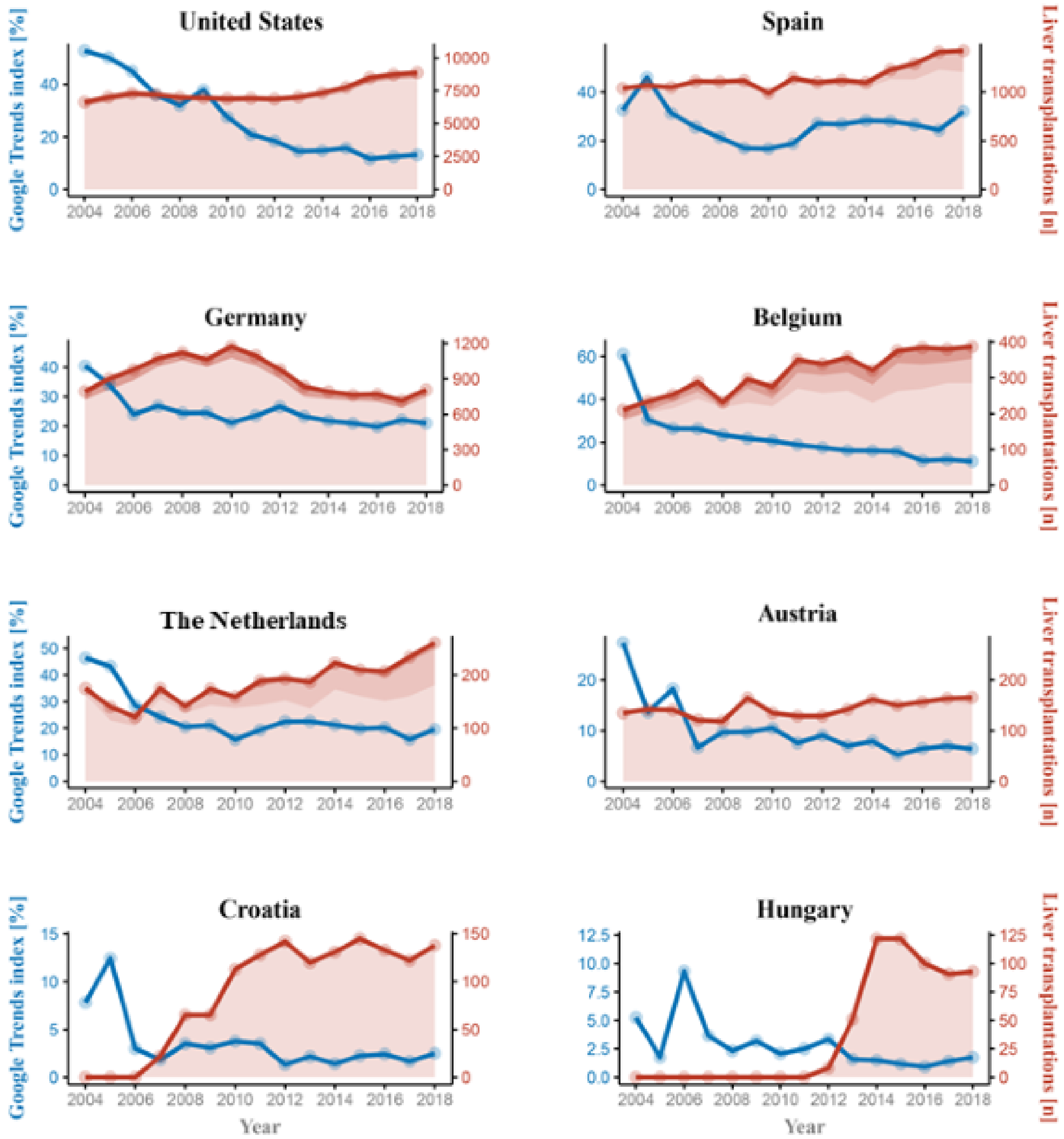


Table 2. The respective year, number of search queries using Google Trends, and the total number of liver transplantations performed in different countries are provided (deceased donor and living donor).

Year	World Google Trends index	United States Google Trends	United States TX[a] total	Spain Google Trends index	Spain TX total	Belgium Google Trends index	Belgium TX total	Luxembourg TX total	The Netherlands Google Trends index	The Netherlands TX total
2004	73.8	59.1667	6642	8.3	1040	22.8	210	1	19.7	175
2005	68.3	60.5833	7015	24.1	1070	11.9	234	2	5.3	140
2006	55.7	49.3333	7302	12.2	1051	16.3	253	5	6.4	121
2007	45.7	45.9167	7202	6.0	1112	8.6	288	1	5.7	175
2008	46.5	45.25	7000	10.3	1108	13.4	232	0	5.5	141
2009	47.5	52.0833	6958	6.3	1119	16.6	296	0	10.8	174
2010	43.3	43.3333	6893	8.9	989	10.4	275	3	8.0	159
2011	41.3	41.8333	6931	7.6	1145	12.5	351	9	8.6	189
2012	40.3	40.0833	6876	13.4	1101	10.1	339	4	11.5	193
2013	37.9	38.1667	7026	13.8	1122	12.3	357	6	8.1	187
2014	36.6	38.0833	7344	13.3	1100	10.3	321	3	9.7	223
2015	39.4	39.9167	7775	17.7	1229	11.3	375	3	9.3	210
2016	39.9	39.75	8497	15.3	1292	11.4	384	3	9.3	207
2017	42.7	43.9167	8740	14.3	1413	9.8	380	9	8.1	234
2018	46.3	50.25	8875	17.3	1426	11.8	387	7	9.0	261

[a] TX: number of transplantations.

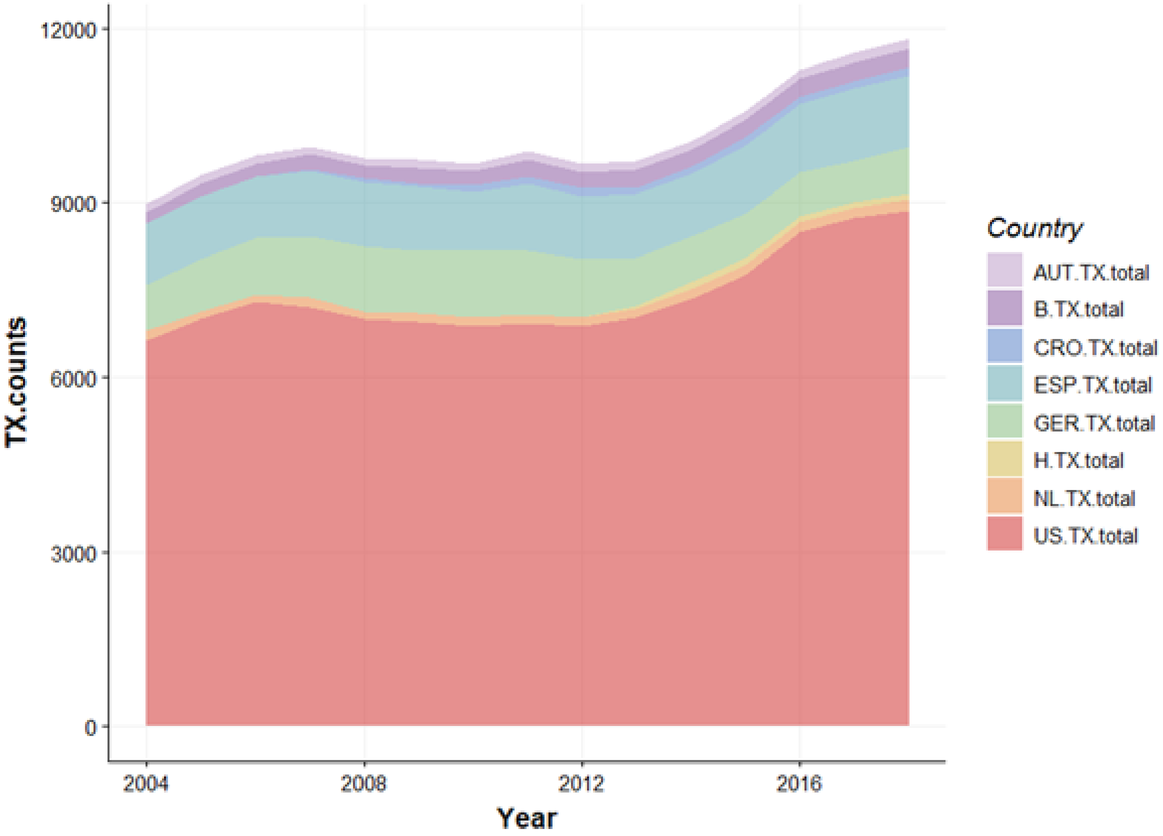
Table 3. The respective year, number of search queries using Google Trends, and the total number of liver transplantations performed are provided (deceased donor and living donor).

Year	Germany Google Trends index	Germany TX[a] total	Austria Google Trends index	Austria TX total	Slovenia TX total	Hungary Google Trends index	Hungary TX total	Croatia Google Trends index	Croatia TX total
2004	61.6	795	15.5	135	24	0.0	0	8.3	0
2005	52.0	901	14.6	142	15	0.0	0	8.3	0
2006	35.3	979	6.9	141	21	8.3	0	0.0	0
2007	32.6	1074	16.2	121	15	1.4	0	0.8	22
2008	35.8	1122	11.7	119	22	1.0	0	1.5	65
2009	35.3	1065	9.4	165	22	2.2	0	2.0	65
2010	29.2	1173	8.8	135	34	0.6	0	1.7	113
2011	33.6	1097	11.8	129	24	2.2	0	1.3	128
2012	36.3	980	12.9	129	38	1.8	8	1.2	142
2013	31.8	836	9.8	142	35	1.3	51	1.9	120
2014	30.1	793	9.3	162	34	1.2	122	1.0	131
2015	27.9	765	5.6	150	43	1.3	122	2.3	145
2016	27.3	771	7.3	157	37	0.8	100	1.3	133
2017	30.3	716	11.2	164	34	1.3	91	1.8	122
2018	27.3	807	6.5	166	29	1.9	93	2.3	138

[a] TX: number of transplantations.

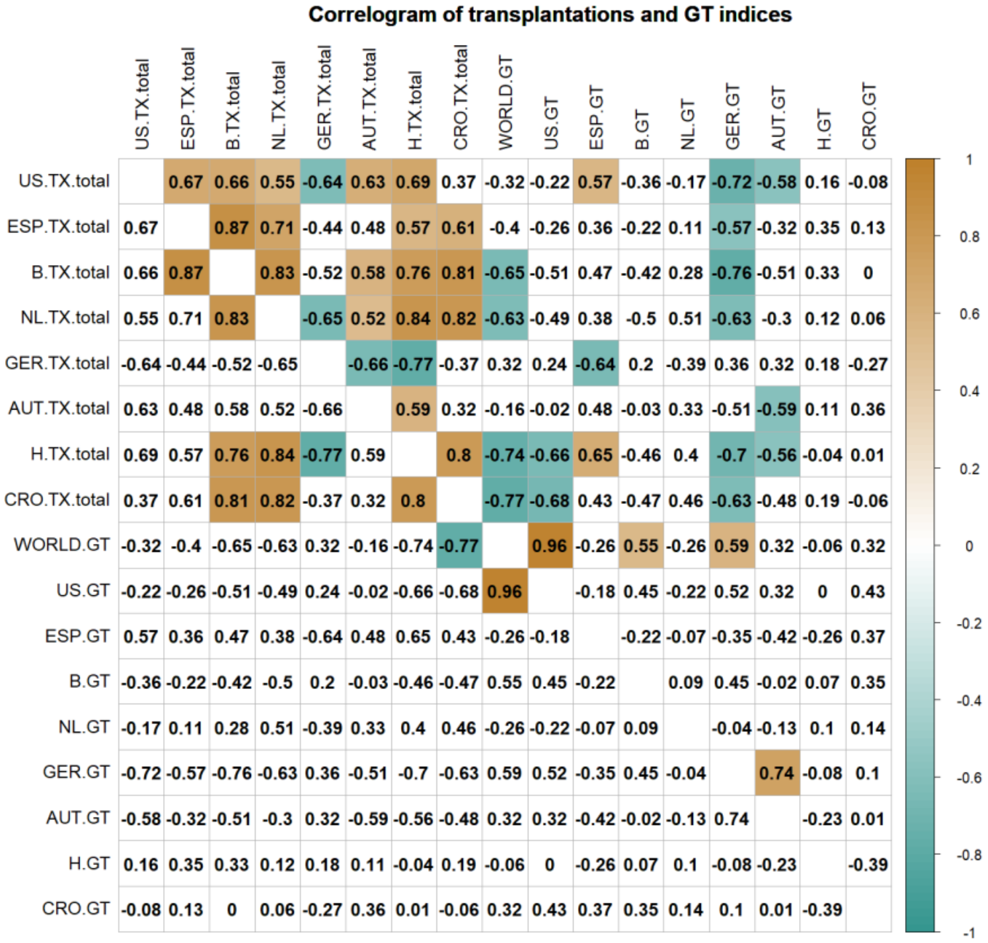
In the same period (2004-2018), UNOS reported the most significant increase in deceased donor liver transplants from 6642 to 8875 (+34.0%). Conversely, the number of living donor donations remained stable during the same period. The number of LTs increased by 24.0% and 18.2% in the Eurotransplant and ONT, respectively (Figure 3; Multimedia Appendix 1), and the number of (both deceased and live donor) LTs increased by 24.0% and 18.2% in the Eurotransplant and ONT, respectively (Figure 3; Multimedia Appendix 1).

Figure 3. Number of liver transplants in Eurotransplant, United Network for Organ Sharing, and Organización Nacional de Trasplantes. AUT: Austria; B: Belgium; CRO: Croatia; ESP: Spain; GER: Germany; GT: Google Trends; H: Hungary; NL: the Netherlands; TX: nuber of transplantations; US: United States.



Belgium and the Netherlands were the only 2 countries in the Eurotransplant region with a mild increase in living donor LT; however, in these countries, a significant decrease in the Google Trends index was observed (Belgium: –48.3%; the Netherlands: –53.3%; Multimedia Appendix 2). Similar downward trends were observed in all Eurotransplant countries. A correlogram of the total transplant numbers and Google Trends indices of the investigated countries are depicted in Figure 4**.** Most notably, even in Croatia, a country with 42 transplantations per million and a dissent solution, the Google Trends index significantly decreased from 7.8 to 2.5 (–75.7%). Google Trends changes and the number of transplants (deceased donor and living donor transplantations) in the respective countries over time are displayed in Figure 2. The number of DCDD donors in the Eurotransplant region and the UNOS area showed a mild increase. In 2018, only 8.08% (145/1795) and 5.71% (537/9412) of deceased donors were DCDD donors for LT in the Eurotransplant and UNOS regions, respectively (Multimedia Appendix 3).

Figure 4. Correlogram of total transplant numbers and Google Trends indices of investigated countries. Correlations are based on the Spearman correlation coefficient. Pairwise correlations between total transplant numbers per country and Google Trends indices were calculated. Significant correlations with *P* values <.05 and <.001 are highlighted by a colorful background in the upper and lower half of the matrix, respectively. AUT: Austria; B: Belgium; CRO: Croatia; ESP: Spain; GER: Germany; GT: Google Trends; H: Hungary; NL: the Netherlands; TX: nuber of transplantations; US: United States.



Correlogram of transplantations and GT indices

	US.TX.total	ESP.TX.total	B.TX.total	NL.TX.total	GER.TX.total	AUT.TX.total	H.TX.total	CRO.TX.total	WORLD.GT	US.GT	ESP.GT	B.GT	NL.GT	GER.GT	AUT.GT	H.GT	CRO.GT
US.TX.total		0.67	0.66	0.55	-0.64	0.63	0.69	0.37	-0.32	-0.22	0.57	-0.36	-0.17	-0.72	-0.58	0.16	-0.08
ESP.TX.total	0.67		0.87	0.71	-0.44	0.48	0.57	0.61	-0.4	-0.26	0.36	-0.22	0.11	-0.57	-0.32	0.35	0.13
B.TX.total	0.66	0.87		0.83	-0.52	0.58	0.76	0.81	-0.65	-0.51	0.47	-0.42	0.28	-0.76	-0.51	0.33	0
NL.TX.total	0.55	0.71	0.83		-0.65	0.52	0.84	0.82	-0.63	-0.49	0.38	-0.5	0.51	-0.63	-0.3	0.12	0.06
GER.TX.total	-0.64	-0.44	-0.52	-0.65		-0.66	-0.77	-0.37	0.32	0.24	-0.64	0.2	-0.39	0.36	0.32	0.18	-0.27
AUT.TX.total	0.63	0.48	0.58	0.52	-0.66		0.59	0.32	-0.16	-0.02	0.48	-0.03	0.33	-0.51	-0.59	0.11	0.36
H.TX.total	0.69	0.57	0.76	0.84	-0.77	0.59		0.8	-0.74	-0.66	0.65	-0.46	0.4	-0.7	-0.56	-0.04	0.01
CRO.TX.total	0.37	0.61	0.81	0.82	-0.37	0.32	0.8		-0.77	-0.68	0.43	-0.47	0.46	-0.63	-0.48	0.19	-0.06
WORLD.GT	-0.32	-0.4	-0.65	-0.63	0.32	-0.16	-0.74	-0.77		0.96	-0.26	0.55	-0.26	0.59	0.32	-0.06	0.32
US.GT	-0.22	-0.26	-0.51	-0.49	0.24	-0.02	-0.66	-0.68	0.96		-0.18	0.45	-0.22	0.52	0.32	0	0.43
ESP.GT	0.57	0.36	0.47	0.38	-0.64	0.48	0.65	0.43	-0.26	-0.18		-0.22	-0.07	-0.35	-0.42	-0.26	0.37
B.GT	-0.36	-0.22	-0.42	-0.5	0.2	-0.03	-0.46	-0.47	0.55	0.45	-0.22		0.09	0.45	-0.02	0.07	0.35
NL.GT	-0.17	0.11	0.28	0.51	-0.39	0.33	0.4	0.46	-0.26	-0.22	-0.07	0.09		-0.04	-0.13	0.1	0.14
GER.GT	-0.72	-0.57	-0.76	-0.63	0.36	-0.51	-0.7	-0.63	0.59	0.52	-0.35	0.45	-0.04		0.74	-0.08	0.1
AUT.GT	-0.58	-0.32	-0.51	-0.3	0.32	-0.59	-0.56	-0.48	0.32	0.32	-0.42	-0.02	-0.13	0.74		-0.23	0.01
H.GT	0.16	0.35	0.33	0.12	0.18	0.11	-0.04	0.19	-0.06	0	-0.26	0.07	0.1	-0.08	-0.23		-0.39
CRO.GT	-0.08	0.13	0	0.06	-0.27	0.36	0.01	-0.06	0.32	0.43	0.37	0.35	0.14	0.1	0.01	-0.39	

Google Trends and Waiting List Mortality

The waiting list mortality did not change significantly in UNOS (–5.2%) and Eurotransplant (–6.2%; Multimedia Appendix 4) regions. Even in Germany, the Eurotransplant country with the highest waiting list mortality (451/1379, 32.7%), the Google Trends index decreased from 40.4 to 21.1 (–48.5%). Furthermore, we analyzed the data of UNOS based on ethnicity. With no significant change over time, Hispanic individuals (2.4%), and American Indians and Alaska Natives (each 4.8%), had a significantly higher mortality on the waiting list than that of all ethnicities (Multimedia Appendix 5). An overview of changes in Google Trends over time, number of transplants (deceased donor, DCDD, and living donor transplantation) in the respective countries is depicted in Figure 3.

Google Trends and LT Program in Spain

Spain exhibited a distinct Google Trends index pattern compared to other countries. The index slightly decreased until 2011 (the year of implementation of the DCDD program in Spain). A campaign for DCDD donors in the public, as well as in hospitals where potential donors are hospitalized, resulted in an increase in the Google Trends index. In the period of 15 years, we could not find a significant decrease in the Google Trends index (–1.8%). In fact, the number of transplanted livers increased because of DCDD by 18.1%. Moreover, there was a decrease in waiting list mortality between 2011 and 2012 (–4.7%). The overall waiting list mortality, too sick to transplant, and the dropout rate for other reasons were also significantly lower in Spain (15%) than in UNOS (31%) or Eurotransplant (29%; Multimedia Appendix 5).

Discussion

Principal Findings

In this study, we found a significant decrease in Google Trends search queries for LT in the UNOS and Eurotransplant regions. As such, public and health care providers' levels of awareness regarding LT are decreasing alarmingly. In Spain, the leading country for transplantation, these findings were not as pronounced. Furthermore, the dropout rate in Spain was significantly lower than that in UNOS and Eurotransplant. Although the need for LT, as the only curative option for chronic liver disease, is increasing, the number of donor organs is also increasing. However, the gap between possible recipients and donors is also increasing. To close this gap, transplant and donor programs, which in part bring awareness to both the public and health care domains, may provide some improvement.

As indicated by the compelling findings presented here, the application of internet data in health care research presents a promising new field. It may further complement and extend the current data sources and foundations [43]. Approximately 90% of US citizens use the internet regularly. According to a data analysis of Pew Research Center (Washington, DC), following an ongoing rapid growth of *going online* and use of social media in the United States over the last decade, it stayed stable over the past 3 years. Comparable data are available in Europe.

Health and *health care* were the number 2 priorities to the US public in 2019. Internet users tend to search for health-related topics accordingly. In fact, more than 80% of all internet users look for health information on the web. Among them, 66% searched for information concerning a specific disease or medical problem (perennially the most popular purpose), and 56% were interested in a certain medical treatment or procedure. After checking emails and using search engines, looking for health topics was the third most frequent activity on the internet. Interestingly, the typical search for health information is on behalf of someone else [44]. The most popular science Facebook group boasts up to 44 million followers [45]. Limited access to internet use, especially internet search for health-related topics, has been found in minorities such as Hispanic, American Indian, and Alaska Native (PEW Research Center). This finding might in part explain the higher mortality and morbidity rates in these ethnicities compared with other ethnicities. Although health care topics on the internet are constantly rising, interest in LT has been decreasing since 2014 all over the world. This trend indicates that the topic LT is underrepresented in the web, despite a small increase seen from 2014 onward. However, the internet (eg, search engines and social media) is the largest platform for awareness programs in the field of liver disease and LT.

To date, very little is published regarding the awareness of LT. This disparity between the low search volumes of the terms relating to LT and the actual increasing number of transplantations may originate in the established low awareness campaigns of LT. Such campaigns are highly useful, as past awareness movements have proved extremely effective. For example, the *Ice Bucket Challenge* promoted awareness of amyotrophic lateral sclerosis. This activity, demonstrated by the dumping of a bucket of ice water over a person's head, went viral in the summer of 2014 and resulted in a nearly 1000-fold increase in the Google Trends index. Subsequently, over US $220 million in funding has been raised worldwide for this rare disease. Several awareness campaigns related to other health issues in recent years have also proved immensely successful. One of the best known includes the *Red Ribbon* movement to fight HIV infection. Even a *World AIDS Day* was initiated on December 1, 1988 [46]. Other programs, such as those promoting the fight against breast cancer, were followed with significant successes in both awareness and funding. An additional notable example is the *Jade Ribbon Campaign*, which was a great success in hepatitis B virus awareness, screening, and physician follow-ups in Chinese Americans. Conversely, the term *liver disease* is strongly underrepresented in the public awareness and, in turn, the World Health Organization's goal to eradicate hepatitis C by 2030 will most likely not be achieved. Even in well-developed countries, there is too little awareness of this disease among health care providers and the broad public [47]. The LT field was even more underrepresented. The reasons for this dearth of awareness are two-fold. The knowledge of primary care providers regarding the possibilities of LT remains insufficient. However, patients complain about a lack of information related to the nature of their disease and the potential to undergo LT.

As shown from past promotion campaigns of various other diseases, public awareness should be the key goal to increase organ donation rates. Spain's case offers evidence of such contention. The overwhelming number of 43.4 donors per million population in the country (2016) reflects the increased level of information provided to the public regarding organ donation. Close attention to the mass media is a key point of the Spanish system and serves a preeminent way to inform the public and raise awareness. As a result of Spain's communication policy, journalists have become extremely important in promoting organ donation. This topic is massive and continuously presented in the media. In 2016, a total of 155 Spanish media reports or news on the topic transplantation were on TV, radio, and printed press releases on the European organ donation day in October. The internet and social media were not included in the survey. The estimated audience comprised 24 million people [48]. Thus, the interest in LT in Spain has remained high over the past 16 years.

In addition, it is important to note that the number of DCDD LTs has increased significantly over the last few years in Spain. This increase in LTs, including DCDD, might reflect the success of awareness campaigns by the ONT. The ONT has established awareness programs across the country, subject to the national Spanish health ministry. Hepatologists and anesthesiologists with special training in the field of LT are representatives of transplantation programs [49]. The fruits of this work were visible in our Google Trends analysis. Specifically, there were increased search rates of the topic, *liver transplantation*, in Spain, alongside increased number of donors, transplantations, and a lower mortality rate on the waiting list. Thus, we conclude that a stable Google Trends index, compared with the global trend, reflects the success story in Spain. This underlines our hypothesis that sensitizing people for the topic could close the gap between supply and demand in LT. Furthermore, the worldwide increase of the search terms *liver transplantation* and *liver transplant* since 2014 may be because of more awareness programs, as well as an increasing number of DCDD and living donor transplants worldwide. Indeed, steps to increase awareness are underway. For example, the first National Patient Advisory Committee of America's Liver Foundation was founded. At present, more than 50 diverse members are trained to raise awareness of the field of LT across the United States. In 2015, legislators were educated about LT and liver disease. Such discussions resulted in an annual Advocacy Day, which allowed for more awareness and an increase in search terms in the United States and worldwide.

The impact of web-based research has grown continuously in the past decade [50]. To date, Google Trends is the only unbiased approach that includes millions of users and has been widely used in economics and health issues. In economics, Google Trends data can help to improve forecasts of the current level of activity for a number of different economic time series such as automobile sales, retail sales, or unemployment [51,52]. Economists have already been at work using Google Trends to make quantitative forecasts [51,53]. Several recent research publications demonstrate that data on web searches from Google Trends can improve the accuracy of forecasts over conventional models. The use of Google data has rapidly spread in the literature to predict other economic indicators, such as analyzing their impact on stock markets and studying bond markets or their impact on commodities [54,55]. Goggle Trends and the field of infodemiology are being widely used in the field of health-related issues as well. Public attention in different fields of health care has been published recently (eg, osteoarthritis, breast cancer, or chronic inflammatory lung disease) [34,56,57]. Furthermore, infodemiology and Google Trends are used to generate awareness profiles and are suitable substitutes for classical data collection, such as surveys [50]. Thus far, Google Trends has been primarily used to monitor disease control and awareness in cancer, HIV, or stroke and also in rare diseases such as antiphospholipid syndrome or systemic lupus erythematosus [35,58-60]. Google Trends offers a wide range of capabilities, with one being the detection of success rates of awareness programs [61,62].

Limitations

Our data indicate multiple novel aspects in the field of LT, such as those concerning donor and recipient awareness. Nonetheless, as with any study, there are some potential limitations. Data should be interpreted with caution in the context of public health and disease awareness. Rationale is 2-fold. First, there was no information about individual searches for the analyzed topics. A bias related to a high number of search queries by health care professionals, industry, or marketing agencies cannot be excluded. Second, it is to some extent elusive which search queries are summarized in the topics defined by Google Trends algorithms, as detailed information on how Google generates these data is not provided. The selection of spelling or terms might affect the results and conclusions; therefore, we chose to use more accurate spelling by native speakers and provide a detailed description of our data-gathering approach to facilitate reproducibility. Misspellings, slang words, or different accent use were considered; foreign languages (eg, English and official languages of neighboring countries) were not taken into consideration. Furthermore, some countries (eg, Hungary and Luxembourg) have a lower number of inhabitants, thus resulting in a small sample size for these countries. This may result in huge variations in Google Trends analyses over time. Another limitation may concern rural areas, as they tend to have limited internet access. Moreover, the internet use of the term *liver transplantation* is low in some countries and their official languages. The importance of accuracy in defining search queries is exemplified by searching Google Trends for the topic *immunosuppressants*. Although not specifically representing LT, immunosuppressants are associated with LT. Hence, using the query *immunosuppressant* may be useful to analyze symptom-related interest but does not sufficiently represent LT awareness. Finally, the number of studies based on Google Trends has been increasing, but so far, there is no standardized procedure for data collection. More guidance by Google is warranted to assist researchers in establishing an optimal search strategy [63].

Conclusions

Google Trends provides a powerful tool for evaluating public interest related to LT and associated liver diseases. According to our study, interest in LT has decreased over the last decade in all investigated countries except Spain. The success story in Spain is encouraging, as it confirms that more awareness campaigns in the field of LT are needed to close the gap between increasing demand and a small supply of potential donor organs. Therefore, international awareness programs are required. In the future, the effects of awareness programs could be evaluated using Google Trends. In line with the goal of higher awareness for solid organ transplantation, Google Trends helps to collect, analyze, report, and disseminate LT-related health data. Google Trends may, therefore, not only drive change and track progress but may also help to improve programs to counteract the current lack of public LT awareness.

Acknowledgments

This work was supported by the Excellence Initiative VASCage (Centre for Promoting Vascular Health in the Ageing Community), a research and development K-Centre (COMET [Competence Centers for Excellent Technologies] program) funded by the Austrian Ministry for Transport, Innovation and Technology, the Austrian Ministry for Digital and Economic Affairs, and the federal states Tyrol, Salzburg, and Vienna.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Liver transplant numbers per country and year.

[PDF File (Adobe PDF File), 134 KB - jmir_v23i8e21656_app1.pdf]

Multimedia Appendix 2

Living, deceased, and donation after circulatory determination of death donors in Eurotransplant by country.

[PDF File (Adobe PDF File), 147 KB - jmir_v23i8e21656_app2.pdf]

Multimedia Appendix 3

Living, deceased, and donation after circulatory determination of death donors in Eurotransplant, United Network for Organ Sharing, and Organización Nacional de Trasplantes.

[PDF File (Adobe PDF File), 128 KB - jmir_v23i8e21656_app3.pdf]

Multimedia Appendix 4

Liver waiting list removal due to death, too sick to transplant, died during transplant, and others.

[PDF File (Adobe PDF File), 161 KB - jmir_v23i8e21656_app4.pdf]

Multimedia Appendix 5

United Network for Organ Sharing data on death removal by ethnicity by year in percentages (%).

[PDF File (Adobe PDF File), 147 KB - jmir_v23i8e21656_app5.pdf]

References

1. Thapar M, Bonkovsky HL. Indications for liver transplant and AASLD guidelines. Hepatology 2015 Jan;61(1):408. [doi: 10.1002/hep.27176] [Medline: 24752818]
2. European Association for the Study of the Liver. EASL Clinical Practice Guidelines: Liver transplantation. J Hepatol 2016 Feb;64(2):433-485. [doi: 10.1016/j.jhep.2015.10.006] [Medline: 26597456]
3. Agopian VG, Petrowsky H, Kaldas FM, Zarrinpar A, Farmer DG, Yersiz H, et al. The evolution of liver transplantation during 3 decades: analysis of 5347 consecutive liver transplants at a single center. Ann Surg 2013 Sep;258(3):409-421. [doi: 10.1097/SLA.0b013e3182a15db4] [Medline: 24022434]
4. Adam R, Karam V, Delvart V, O'Grady J, Mirza D, Klempnauer J, All contributing centers (www.eltr.org), European LiverIntestine Transplant Association (ELITA). Evolution of indications and results of liver transplantation in Europe. A report from the European Liver Transplant Registry (ELTR). J Hepatol 2012 Sep;57(3):675-688 [FREE Full text] [doi: 10.1016/j.jhep.2012.04.015] [Medline: 22609307]
5. Dutkowski P, De Rougemont O, Müllhaupt B, Clavien P. Current and future trends in liver transplantation in Europe. Gastroenterology 2010 Mar;138(3):802-809. [doi: 10.1053/j.gastro.2010.01.030] [Medline: 20096694]
6. Dutkowski P, Linecker M, DeOliveira ML, Müllhaupt B, Clavien P. Challenges to liver transplantation and strategies to improve outcomes. Gastroenterology 2015 Feb;148(2):307-323. [doi: 10.1053/j.gastro.2014.08.045] [Medline: 25224524]
7. Kwong A, Kim WR, Lake JR, Smith JM, Schladt DP, Skeans MA, et al. OPTN/SRTR 2018 Annual Data Report: Liver. Am J Transplant 2020 Jan;20 Suppl s1:193-299. [doi: 10.1111/ajt.15674] [Medline: 31898413]
8. Fayek SA, Quintini C, Chavin KD, Marsh CL. The current state of liver transplantation in the United States: perspective from American Society of Transplant Surgeons (ASTS) Scientific Studies Committee and Endorsed by ASTS Council. Am J Transplant 2016 Nov;16(11):3093-3104 [FREE Full text] [doi: 10.1111/ajt.14017] [Medline: 27545282]
9. Samuel D, Coilly A. Management of patients with liver diseases on the waiting list for transplantation: a major impact to the success of liver transplantation. BMC Med 2018 Aug 01;16(1):113 [FREE Full text] [doi: 10.1186/s12916-018-1110-y] [Medline: 30064414]
10. Müllhaupt B, Dimitroulis D, Gerlach JT, Clavien P. Hot topics in liver transplantation: organ allocation - extended criteria donor - living donor liver transplantation. J Hepatol 2008;48 Suppl 1:58-67 [FREE Full text] [doi: 10.1016/j.jhep.2008.01.013] [Medline: 18308415]
11. Boteon YL, Laing RW, Schlegel A, Wallace L, Smith A, Attard J, et al. Combined hypothermic and normothermic machine perfusion improves functional recovery of extended criteria donor livers. Liver Transpl 2018 Dec;24(12):1699-1715 [FREE Full text] [doi: 10.1002/lt.25315] [Medline: 30058119]
12. Giorgakis E, Khorsandi SE, Jassem W, Heaton N. DCD consensus and futility in liver transplantation. J Hepatol 2018 Jul;69(1):255-256. [doi: 10.1016/j.jhep.2018.03.032] [Medline: 29655976]
13. Eysenbach G. Infodemiology: The epidemiology of (mis)information. Am J Med 2002 Dec 15;113(9):763-765. [doi: 10.1016/s0002-9343(02)01473-0] [Medline: 12517369]
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;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
15. Eysenbach G. How to fight an infodemic: the four pillars of infodemic management. J Med Internet Res 2020 Jun 29;22(6):e21820 [FREE Full text] [doi: 10.2196/21820] [Medline: 32589589]
16. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. Am J Prev Med 2011 May;40(5 Suppl 2):154-158. [doi: 10.1016/j.amepre.2011.02.006] [Medline: 21521589]
17. 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]

18. 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]
19. Adrover C, Bodnar T, Huang Z, Telenti A, Salathé M. Identifying adverse effects of HIV drug treatment and associated sentiments using Twitter. JMIR Public Health Surveill 2015;1(2):e7 [FREE Full text] [doi: 10.2196/publichealth.4488] [Medline: 27227141]
20. Zheluk A, Quinn C, Hercz D, Gillespie JA. Internet search patterns of human immunodeficiency virus and the digital divide in the Russian Federation: infoveillance study. J Med Internet Res 2013;15(11):e256 [FREE Full text] [doi: 10.2196/jmir.2936] [Medline: 24220250]
21. Mavragani A, Ochoa G. Forecasting AIDS prevalence in the United States using online search traffic data. J Big Data 2018 May 19;5(1):5. [doi: 10.1186/s40537-018-0126-7]
22. Koschack J, Weibezahl L, Friede T, Himmel W, Makedonski P, Grabowski J. Scientific versus experiential evidence: discourse analysis of the chronic cerebrospinal venous insufficiency debate in a multiple sclerosis forum. J Med Internet Res 2015 Jul 01;17(7):e159 [FREE Full text] [doi: 10.2196/jmir.4103] [Medline: 26133525]
23. Risson V, Saini D, Bonzani I, Huisman A, Olson M. Patterns of treatment switching in multiple sclerosis therapies in US patients active on social media: application of social media content analysis to health outcomes research. J Med Internet Res 2016;18(3):e62 [FREE Full text] [doi: 10.2196/jmir.5409] [Medline: 26987964]
24. Rose SW, Jo CL, Binns S, Buenger M, Emery S, Ribisl KM. Perceptions of menthol cigarettes among Twitter users: content and sentiment analysis. J Med Internet Res 2017 Feb 27;19(2):e56 [FREE Full text] [doi: 10.2196/jmir.5694] [Medline: 28242592]
25. 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]
26. Yom-Tov E, Gabrilovich E. Postmarket drug surveillance without trial costs: discovery of adverse drug reactions through large-scale analysis of web search queries. J Med Internet Res 2013;15(6):e124 [FREE Full text] [doi: 10.2196/jmir.2614] [Medline: 23778053]
27. Santos JC, Matos S. Analysing Twitter and web queries for flu trend prediction. Theor Biol Med Model 2014 May 07;11 Suppl 1:S6 [FREE Full text] [doi: 10.1186/1742-4682-11-S1-S6] [Medline: 25077431]
28. Effenberger M, Kronbichler A, Shin JI, Mayer G, Tilg H, Perco P. Association of the COVID-19 pandemic with internet search volumes: a Google Trends Analysis. Int J Infect Dis 2020 Jun;95:192-197 [FREE Full text] [doi: 10.1016/j.ijid.2020.04.033] [Medline: 32305520]
29. Nsoesie EO, Cesare N, Müller M, Ozonoff A. COVID-19 Misinformation spread in eight countries: exponential growth modeling study. J Med Internet Res 2020 Dec 15;22(12):e24425 [FREE Full text] [doi: 10.2196/24425] [Medline: 33264102]
30. Mavragani A. Tracking COVID-19 in Europe: Infodemiology approach. JMIR Public Health Surveill 2020 Apr 20;6(2):e18941 [FREE Full text] [doi: 10.2196/18941] [Medline: 32250957]
31. Jacobson NC, Lekkas D, Price G, Heinz MV, 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. [doi: 10.2196/19347] [Medline: 32459186]
32. 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]
33. Arora VS, McKee M, Stuckler D. Google Trends: Opportunities and limitations in health and health policy research. Health Policy 2019 Mar;123(3):338-341. [doi: 10.1016/j.healthpol.2019.01.001] [Medline: 30660346]
34. Kaleem T, Malouff TD, Stross WC, Waddle MR, Miller DH, Seymour AL, 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]
35. Ling R, Lee J. Disease monitoring and health campaign evaluation using Google search activities for HIV and aids, stroke, colorectal cancer, and marijuana use in Canada: a retrospective observational study. JMIR Public Health Surveill 2016 Oct 12;2(2):e156 [FREE Full text] [doi: 10.2196/publichealth.6504] [Medline: 27733330]
36. Kronbichler A, Effenberger M, Shin JI, Koppelstätter C, Denicolò S, Rudnicki M, et al. Is there decreasing public interest in renal transplantation? A Google Trends Analysis. J Clin Med 2020 Apr 07;9(4):1048 [FREE Full text] [doi: 10.3390/jcm9041048] [Medline: 32272783]
37. Van de Belt TH, Engelen LJ, Berben SA, Teerenstra S, Samsom M, Schoonhoven L. Internet and social media for health-related information and communication in health care: preferences of the Dutch general population. J Med Internet Res 2013;15(10):e220 [FREE Full text] [doi: 10.2196/jmir.2607] [Medline: 24088272]
38. Ybarra ML, Suman M. Help seeking behavior and the internet: a national survey. Int J Med Inform 2006 Jan;75(1):29-41. [doi: 10.1016/j.ijmedinf.2005.07.029] [Medline: 16129659]
39. United Network for Organ Sharing (UNOS). URL: https://unos.org [accessed 2021-07-23]
40. National Transplant Organization (ONT). URL: http://www.ont.es [accessed 2021-07-23]
41. Eurotransplant. URL: https://www.eurotransplant.org [accessed 2021-07-23]

42. Google Trends. URL: https://trends.google.com/trends [accessed 2021-07-23]
43. Mavragani A, Ochoa G. Google Trends in infodemiology and infoveillance: methodology framework. JMIR Public Health Surveill 2019 May 29;5(2):e13439 [FREE Full text] [doi: 10.2196/13439] [Medline: 31144671]
44. Gidwani R, Zulman D. Association between acute medical exacerbations and consuming or producing web-based health information: analysis from Pew Survey Data. J Med Internet Res 2015 Jun 23;17(6):e145 [FREE Full text] [doi: 10.2196/jmir.3801] [Medline: 26104000]
45. Li J, Theng Y, Foo S. Predictors of online health information seeking behavior: changes between 2002 and 2012. Health Informatics J 2015 Aug 10:804-814. [doi: 10.1177/1460458215595851] [Medline: 26261218]
46. No authors listed. World AIDS Day - December 1, 2019. MMWR Morb Mortal Wkly Rep 2019 Nov 29;68(47):1089 [FREE Full text] [doi: 10.15585/mmwr.mm6847a1] [Medline: 31851652]
47. Schaefer B, Viveiros A, Al-Zoairy R, Blach S, Brandon S, Razavi H, et al. Disease burden of hepatitis C in the Austrian state of Tyrol - Epidemiological data and model analysis to achieve elimination by 2030. PLoS One 2018;13(7):e0200750 [FREE Full text] [doi: 10.1371/journal.pone.0200750] [Medline: 30001427]
48. Matesanz R, Marazuela R, Coll E, Mahíllo B, Domínguez-Gil B. About the opt-out system, live transplantation, and information to the public on organ donation in Spain … Y olé!. Am J Transplant 2017 Jun;17(6):1695-1696 [FREE Full text] [doi: 10.1111/ajt.14296] [Medline: 28382792]
49. de la Rosa G, Fondevila C, Navasa M. Liver transplantation in Spain. Liver Transpl 2016 Sep;22(9):1259-1264 [FREE Full text] [doi: 10.1002/lt.24484] [Medline: 27197947]
50. Jun S, Yoo HS, Choi S. Ten years of research change using Google Trends: from the perspective of big data utilizations and applications. Technol Forecast Soc Change 2018 May;130:69-87. [doi: 10.1016/j.techfore.2017.11.009]
51. Fritzsch B, Wenger K, Sibbertsen P, Ullmann G. Can Google Trends improve sales forecasts on a product level? Appl Econ Lett 2019 Nov 01;27(17):1409-1414. [doi: 10.1080/13504851.2019.1686110]
52. Nagao S, Takeda F, Tanaka R. Nowcasting of the U.S. unemployment rate using Google Trends. Finance Res Lett 2019 Sep;30:103-109. [doi: 10.1016/j.frl.2019.04.005]
53. Bulut L. Google Trends and the forecasting performance of exchange rate models. J Forecast 2017 Nov 29;37(3):303-315. [doi: 10.1002/for.2500]
54. Peri M, Baldi L, Vandone D. Price discovery in commodity markets. Appl Econ Lett 2013 Mar;20(4):397-403. [doi: 10.1080/13504851.2012.709590]
55. Ben-Rephael A, Da Z, Israelsen R. It depends on where you search: institutional investor attention and underreaction to news. Review of financial studies. 2017 SEP 2017;30(9):3009-3047. [doi: 10.1093/rfs/hhx031]
56. Jellison SS, Bibens M, Checketts J, Vassar M. Using Google Trends to assess global public interest in osteoarthritis. Rheumatol Int 2018 Dec;38(11):2133-2136. [doi: 10.1007/s00296-018-4158-2] [Medline: 30218145]
57. Boehm A, Pizzini A, Sonnweber T, Loeffler-Ragg J, Lamina C, Weiss G, et al. Using Google Trends to investigate global COPD awareness. Eur Respir J 2019 Aug;54(2):1901339. [doi: 10.1183/13993003.01339-2019] [Medline: 31395673]
58. Mahroum N, Bragazzi NL, Brigo F, Waknin R, Sharif K, Mahagna H, et al. Capturing public interest toward new tools for controlling human immunodeficiency virus (HIV) infection exploiting data from Google Trends. Health Informatics J 2019 Dec 11;25(4):1383-1397 [FREE Full text] [doi: 10.1177/1460458218766573] [Medline: 29638172]
59. Sciascia S, Radin M, Unlu O, Erkan D, Roccatello D. Infodemiology of antiphospholipid syndrome: merging informatics and epidemiology. Eur J Rheumatol 2018 Jul;5(2):92-95 [FREE Full text] [doi: 10.5152/eurjrheum.2018.17105] [Medline: 30185355]
60. Sciascia S, Radin M. What can Google and Wikipedia can tell us about a disease? Big Data trends analysis in Systemic Lupus Erythematosus. Int J Med Inform 2017 Nov;107:65-69. [doi: 10.1016/j.ijmedinf.2017.09.002] [Medline: 29029693]
61. McLean S, Lennon P, Glare P. Internet search query analysis can be used to demonstrate the rapidly increasing public awareness of palliative care in the USA. BMJ Support Palliat Care 2019 Mar;9(1):40-44. [doi: 10.1136/bmjspcare-2016-001171] [Medline: 28130324]
62. Patel MS, Halpern JA, Desai AS, Keeter MK, Bennett NE, Brannigan RE. Success of prostate and testicular cancer awareness campaigns compared to breast cancer awareness month according to internet search volumes: a Google Trends Analysis. Urology 2020 May;139:64-70. [doi: 10.1016/j.urology.2019.11.062] [Medline: 32001306]
63. 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;9(10):e109583 [FREE Full text] [doi: 10.1371/journal.pone.0109583] [Medline: 25337815]

Abbreviations

DCDD: donation after circulatory determination of death
LT: liver transplantation
ONT: Organización Nacional de Trasplantes
UNOS: United Network for Organ Sharing

Edited by R Kukafka; submitted 20.06.20; peer-reviewed by A Mavragani, T Cruvinel; comments to author 29.08.20; revised version received 23.10.20; accepted 21.06.21; published 17.08.21.

Please cite as:
Effenberger M, Kronbichler A, Bettac E, Grabherr F, Grander C, Adolph TE, Mayer G, Zoller H, Perco P, Tilg H
Using Infodemiology Metrics to Assess Public Interest in Liver Transplantation: Google Trends Analysis
J Med Internet Res 2021;23(8):e21656
URL: https://www.jmir.org/2021/8/e21656
doi:10.2196/21656
PMID:34402801

©Maria Effenberger, Andreas Kronbichler, Erica Bettac, Felix Grabherr, Christoph Grander, Timon Erik Adolph, Gert Mayer, Heinz Zoller, Paul Perco, Herbert Tilg. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 17.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Development of a COVID-19 Web Information Transmission Structure Based on a Quadruple Helix Model: Webometric Network Approach Using Bing

Yu Peng Zhu[1,2], PhD; Han Woo Park[2,3,4], PhD

[1]Blockchain Policy Research Center, Cyber Emotions Research Institute, Yeungnam University, Gyeongsan-si, Republic of Korea
[2]Department of Media and Communication, Yeungnam University, Gyeongsan-si, Republic of Korea
[3]Interdisciplinary Graduate Programs of Digital Convergence Business, Yeungnam University, Gyeongsan-si, Republic of Korea
[4]Interdisciplinary Graduate Programs of East Asian Cultural Studies, Yeungnam University, Gyeongsan-si, Republic of Korea

Corresponding Author:
Han Woo Park, PhD
Department of Media and Communication
Yeungnam University
214-1 Dae-dong
Gyeongsan-si
Republic of Korea
Phone: 82 53 810 2275
Email: hanpark@ynu.ac.kr

Abstract

Background: Developing an understanding of the social structure and phenomenon of pandemic information sources worldwide is immensely significant.

Objective: Based on the quadruple helix model, the aim of this study was to construct and analyze the structure and content of the internet information sources regarding the COVID-19 pandemic, considering time and space. The broader goal was to determine the status and limitations of web information transmission and online communication structure during public health emergencies.

Methods: By sorting the second top-level domain, we divided the structure of network information sources into four levels: government, educational organizations, companies, and nonprofit organizations. We analyzed the structure of information sources and the evolution of information content at each stage using quadruple helix and network analysis methods.

Results: The results of the structural analysis indicated that the online sources of information in Asia were more diverse than those in other regions in February 2020. As the pandemic spread in April, the information sources in non-Asian regions began to diversify, and the information source structure diversified further in July. With the spread of the pandemic, for an increasing number of countries, not only the government authorities of high concern but also commercial and educational organizations began to produce and provide significant amounts of information and advice. Nonprofit organizations also produced information, but to a lesser extent. The impact of the virus spread from the initial public level of the government to many levels within society. After April, the government's role in the COVID-19 network information was central. The results of the content analysis showed that there was an increased focus on discussion regarding public health–related campaign materials at all stages. The information content changed with the changing stages. In the early stages, the basic situation regarding the virus and its impact on health attracted most of the attention. Later, the content was more focused on prevention. The business and policy environment also changed from the beginning of the pandemic, and the social changes caused by the pandemic became a popular discussion topic.

Conclusions: For public health emergencies, some online and offline information sources may not be sufficient. Diversified institutions must pay attention to public health emergencies and actively respond to multihelical information sources. In terms of published messages, the educational sector plays an important role in public health events. However, educational institutions release less information than governments and businesses. This study proposes that the quadruple helix not only has research significance in the field of scientific cooperation but could also be used to perform effective research regarding web information during crises. This is significant for further development of the quadruple helix model in the medical internet research area.

(J Med Internet Res 2021;23(8):e27681) doi:10.2196/27681

KEYWORDS

quadruple helix model; COVID-19; structural analysis; content analysis; network analysis; public health; webometrics; infodemiology; infoveillance; development; internet; online health information; structure; communication; big data

Introduction

Background

Since the first reported case of COVID-19 in late 2019, the disease rapidly spread to become a pandemic in March 2020. An infectious disease caused by a pathogen generally spreads to a living host, and it is easily transferable from the infected. As of January 7, 2021, over 87 million people have been infected with the virus in 190 countries and regions worldwide since the outbreak of COVID-19 in February 2020, resulting in a global catastrophe [1].

Social disasters, including infectious diseases, must be controlled through a process of actual data–based analysis [2]. Real-time assessment is critical for disaster monitoring; special attention must be paid to rapid analysis using relevant social and cultural data at both the macro and micro scales. The core of sociocultural data analysis is to understand, identify, and even predict the risk of transmission. Thus, a system needs to be constructed to collect data based on the disaster type or area, and to support the spontaneous decision-making process. Without this system, the public would face an information overload, as they would feel burdened dealing with an enormous amount of information in critical situations [3].

In today’s knowledge-based society, information production goes beyond traditional media organizations and involves many entities. In particular, with the increased popularity of smartphones, the amount of online information produced by individuals and organizations has grown exponentially. As the information production process becomes increasingly complex, traditional media companies face a situation in which the so-called legacy media’s usage time has decreased and is now competing with various sources [4]. Search engines and web portals are catering to a wider range of user needs than traditional alternatives [5]. Low cost is an important reason that internet information channels have more advantages than traditional information channels [6]. This phenomenon is especially obvious when significant events occur. The same is true for the information about COVID-19. Fear, anger, and other emotions also lead people to believe and spread online information available through nontraditional media, regardless of whether it is fake [7]. This implies that not only individual media and informal organizations but also a large number of formal organizations such as governments, academic institutions, and formal public organizations have started using online media to produce and disseminate information. However, sensitivity to fake news is also often influenced by political ideology [8]. In some countries, citizens do not support direct government control of the news, and are more concerned with information sources in cooperation with the media and other nongovernmental organizations [9]. Therefore, web media represent a strong competitor to traditional media in terms of both production and services.

COVID-19 has had a strong impact on the media system [10-12]. People have created significant numbers of online documents by utilizing new media sources such as Facebook and Twitter [13-15]. In fact, governments in some countries have used these forms of new media to build platforms to help combat the virus [16]. For example, due to the COVID-19 outbreak in Wuhan in February 2020, and the shortage of medical resources and services, Weibo, one of the largest new media companies in China, cooperated with the local government to set up a citizen assistance platform on which citizens with real-name identification could ask for help. This new mode of interaction is difficult to create with traditional media. To deal with the COVID-19 pandemic that is currently threatening the world, it is evident that the greater the number of internet information sources, the stronger the social immune system.

Therefore, in this study, we collected information sources with high online presence through big-data techniques in countries with confirmed COVID-19 cases. Additionally, the information related to COVID-19 in three stages (from February to April to July 2020) was analyzed in detail, and the morphology and trend of the web big data in these first 6 months of the pandemic are discussed. From these large-scale online big data, the information dissemination trends of educational institutes, enterprises, and government, and their contents were analyzed. In the case of an emergency or a disaster, this study can systematically explain the multielement spiral structure of the information transmission source in the cyberspace of major countries, which has high academic and social value.

In general, based on the quadruple helix and network analysis method, this study constructed and analyzed the structure and content of internet information sources of COVID-19 considering time and space. The aim was to determine the status and limitations of web information transmission and online communication structure in public health emergencies. Moreover, based on the content revealed, valuable suggestions are proposed to contribute to the internet communication of future public health events.

Online Information Sources

With the rapid growth of the internet, web data analysis (often called “webometrics”) has become important, and its quantitative vastness and content diversity have been increasing accordingly. As mobile phones, tablets, and other mobile terminals have been growing in popularity, people usually use these terminals to obtain information instead of traditional media. Although some mainstream media outlets have their own web feeds, people generally use digital feeds from search engines such as Bing or Google to obtain information. Therefore, the study and analysis of comprehensive network information is often more objective than the study of specific news media, and more comprehensive information can be collated. For example, Thelwall [17] used Wikipedia data collected by Bing to study public interest in astronomy. Park et al [18] used Twitter and YouTube to analyze the spread of the Occupy Wall Street movement. Park and Lim [19] analyzed North Korean propaganda changes using YouTube media data. Cho and Park [16] used network activity information about the agriculture, forestry, and fishery departments to discuss the use of internet innovation by government organizations. This literature indicates that analysis and research on web sources have drawn useful conclusions in many areas to date.

There has also been substantial research regarding online sources of information at the time a disaster occurs. Jung and Park [20] used webometric methods to track and analyze the information networks of various organizations during the Gumi chemical spill in South Korea. They found that the flow of information between agencies had an impact on mobilizing emergency facilities and planning specific emergency responses. Online sources of information can also help alleviate the damage caused by disasters. Allaire [21] studied the Bangkok floods and found that social media users were obtaining real-time updates that could help to reduce their losses. Park [22] analyzed YouTube social activities during the 2016 South Korean earthquake, and found that YouTube became a channel to raise public crisis awareness and promote safety strategies. Kim et al [23] also elaborated on the role of social media in relaying information during disasters. By studying data from online information sources during the 2017 storms in the United States, it was found that the flow of information across the network was controlled by many types of users. Song et al [24] studied the differences and the range of emotions people felt toward local online channels, including publishing boards, Twitter, cafes, blogs, and news, that delivered information related to MERS (Middle East Respiratory Syndrome) [24]. Some scholars also analyzed the network information source data regarding COVID-19. For example, Park et al [25] collected Twitter data and found that monitoring public dialog and rapidly spreading media news can help professionals make complex and rapid decisions. However, web page data in internet information sources are often more stable than those in a social media environment [26]; therefore, we adopted web page data for analysis.

To grasp and respond to the situation of worldwide disasters such as COVID-19, we collected and analyzed the network information source data of countries with a large number of confirmed cases to understand how the disaster-related information provided by multiple sources changes over time. We identified three time periods (February, April, and July 2020), and performed a detailed analysis of the differences in information sources at these times. We also studied the structure of national cyberspace information sources. Through these analyses, we identified the structural evolution of web information publishers and clearly revealed the dynamic changes of information content in each period.

Research Questions

Our primary research questions were as follows: (1) What are the structures and form of the COVID-19 web-mediated network among countries? (2) Are there differences in the keywords and topics of COVID-19–related online information at different stages?

Methods

Data Collection

Data were collected using Webometric Analyst 4.1 through Bing, which is one of the most widely used search engines that is available in most countries and regions, including Mainland China. In addition to Google, the Bing search engine is also often used to carry out scientific research [27-29]. Although Google is the world's largest search engine, it is not available in some regions, including mainland China. Since web page data from China were very important for this study, we used the Bing platform for data collection. Website and domains obtained through the search application programming interface service of Bing were analyzed. In February 2020, the most widely used COVID-19 keyword in the world was "coronavirus." Therefore, the keyword for the data collected on February 12 was "coronavirus." After February, the terms "COVID-19" and "2019-nCOV" were also widely used. Therefore, we chose "COVID-19 OR Coronavirus OR 2019-nCOV" for the keyword searches performed on April 17 and July 22, 2020. Data were collected in real time, instead of collecting all information during a certain period. In other words, the data for these three time points (February 12, April 17, and July 22, 2020) are the results of real-time relevant searches on Bing for that day. Instant messages were not limited to the time of publication, and they may contain previously published information that is still highly popular. In addition, instant messages can reflect the actual state of the internet data at that time.

The time of the first collection was in the initial stage of the outbreak, the time of the second collection corresponded to the stage at which the number of new diagnoses had leveled off after spreading worldwide, and the time of the third collection corresponded to the stage when the number of newly diagnosed patients increased sharply as a second wave. For February, we collected data for all 28 countries and regions with confirmed cases, with a total of 9149 data points. For April 17, we collected data from 29 countries and regions with over 7000 confirmed cases and obtained 14,768 data points. For July 22, we collected data of over 70,000 people who received a diagnosis in 30 countries and regions, and obtained 14,483 pieces of data. Our data were collected from the top countries with the highest number of diagnoses per stage, not only from English-speaking countries. To ensure consistency, we only analyzed the English data on the webpage.

Quadruple Helix

Quadruple helix is a research method based on triple helix, which was proposed in 1995 [30]. Researchers studied the development of the knowledge-based economic structure through the spiral relationships among universities, industries, and the government. At first, the triple helix model was used to explain the interaction among academia, government, and industries, and was often used in research related to knowledge production [31]. However, with the development of the triple helix theory, more elements were considered. In 2009, Carayannis and Campbell [32] introduced elements representing the public into the spiral model, such as the civil society and media, thereby forming a quadruple helix model; they added a research method at the level of new technologies and social needs. In 2010, Carayannis and Campbell [33] added the natural environment factor and constructed the quintuple helix model. Based on this factor, the relationship between innovation and sustainable development can be discussed. During cooperation and communication between various organizations, when one kind of organization occupies a dominant position, it can be considered that this organization is separated from the collection of various organizations, and the relationships among different organizations can be studied and explained through the quadruple, quintuple, or n-tuple helix concept [31]. This spiral structure does not always exist only in the academic, government, and industrial dimensions.

We collated the second top-level domain (TLD) data, which were categorized as data from commercial organizations, educational institutes, governments, and nonprofit organizations. A total of 38,399 domains were collected. To better classify the effective levels, we first sorted all of the collected second TLD data. After frequency analysis, the second TLD data that had an occurrence frequency higher than 1% (91/9149; 148/14,767; 145/14,483) were selected for classification. From 15,813 data units, we extracted four levels, namely governments, commercial enterprises, educational institutes, and nonprofit organizations. We used the quadruple helix model to analyze the structural dynamics of the four institutional levels at different periods in detail. For convenience in figures, we abbreviate government domains such as “.gov,” “gob,” and “.go” collectively as “G” (governments); educational domains such as “.edu” and “.ac” as “E” (educational institutes); commercial domains such as “.com” and “.co” as “C” (commercial organizations); and “.org” and “.or” nonprofit domains as “O” (nonprofit organizations).

Network Analysis

A network analysis method was used to analyze the structure of the quadruple helix in detail. Network analysis is a method of quantitative analysis of nodes and connections in a network. When individuals and organizations act as nodes, the connection between them acts as a link. Through the quantitative results of the structure, the characteristics and nature of the network composed of these entities can be analyzed [34]. Network analysis has been widely used in social science research such as in social media use, knowledge dissemination, and organizational cooperation [35,36]. Although nodes have the same properties in a one-mode network, nodes differ in two-mode networks. Thus, in the one-mode network, nodes are the institutional components of a quadruple helix: government, private/business, educational institutions, and nonprofit organizations. By contrast, the two-mode network focuses on the relationship between the analyzed countries and the four institutional types. Centrality indices are important quantitative indices in network analysis, including degree centrality, betweenness centrality, eigenvector centrality, closeness centrality, and others [37-40]. The centrality index used in this study was degree centrality. Degree indicates the direct relationship between the nodes [39,40]. In this study, the degree was mainly used to determine whether the governments, education institutions, nonprofit organizations, and commercial organizations of different countries have similar information, and to observe the helix degree of different countries and the four fields. Although betweenness centrality is an important indicator for evaluating the influence of a mediating effect, the link between countries in this study is common information without mediating phenomena; thus, it was not used for this analysis. In addition, the eigenvector centrality is an index to evaluate the importance of each node connected to other nodes. However, in this study, based on the importance of government, education, public authorities, and companies, it was considered to be less important to evaluate the significance of the connected countries’ eigenvector centrality. Closeness centrality is an indicator of the shortest distance, and was also considered to be of little significance for this study. However, the degree centrality index can judge the strength of direct connections between countries and domains. In other words, the more countries connected in a field, the stronger the influence of this field. Therefore, only degree centrality was used for this analysis.

We used UCINET6 for network analysis and network visualization, including triple helix network analysis and content analysis. In the content analysis, we used the convergence of iterated correlations (CONCOR) method to cluster the semantic network in which words are regarded as nodes and cooccurrence between words forms a tie. CONCOR is a method of performing repeated cross-node correlation analyses to identify the appropriate level of similarity [41]. In other words, we first organized the relationships among words into matrices, thus forming a network of relationships among words. We then calculated the correlation coefficients between the rows and columns in the matrix and carried out the same calculation for the obtained correlation coefficient matrix. After repeated calculations, a correlation coefficient matrix consisting of only 1 and –1 was obtained, which was thus divided into two categories. We then performed the same calculation for both categories again and obtained four different clusters.

Results

The hit counts and domains for each country (or region) are compiled and listed in Tables 1-3 for the three time periods, respectively. We standardized the number of hits and domain names for the three months (February, April, and July), with the maximum value set to 100 and the rest being the ratio of the original value to the maximum value multiplied by 100.

Table 1. Hit counts and domains for February (N=28).

Country	Hit counts	Domains
Australia	100.00	100.00
Canada	32.46	53.23
Italy	24.69	67.93
United Kingdom	23.93	33.08
Germany	21.05	43.98
France	18.30	33.46
Spain	13.66	44.61
Belgium	10.41	38.91
Japan	7.29	81.62
India	5.99	33.33
Mainland China	5.61	47.40
Singapore	5.34	69.84
Malaysia	4.10	28.77
United States	3.11	64.39
Hong Kong (China)	2.26	68.44
United Arab Emirates	2.22	7.73
South Korea	1.63	32.07
Taiwan	1.29	35.61
Philippines	1.29	29.91
Sweden	1.10	58.17
Sri Lanka	0.83	10.65
Vietnam	0.70	40.18
Finland	0.61	19.90
Russia	0.42	47.02
Thailand	0.18	41.70
Macau	0.16	14.45
Nepal	0.06	5.83
Cambodia	0.02	7.35

Table 2. Hit counts and domains for April (N=29).

Country	Hit counts	Domains
Canada	100.00	81.26
France	69.96	79.83
England	61.98	58.84
Germany	53.23	83.51
Brazil	51.71	88.14
Italy	46.01	75.92
US	32.78	99.88
Sweden	31.67	100.00
Belgium	30.61	44.13
Japan	30.42	61.80
Austria	21.48	41.04
Netherlands	20.95	93.24
Spain	18.17	59.43
Turkey	18.10	49.23
Chile	16.73	55.63
India	16.35	43.06
Switzerland	13.92	67.50
Denmark	13.50	84.82
Russia	9.16	27.05
Portugal	9.13	54.09
Ireland	8.33	45.20
Korea	6.39	46.74
Peru	5.86	31.55
Mainland China	4.18	35.11
Poland	3.84	75.56
Romania	3.31	49.11
Ecuador	2.95	41.28
Israel	2.67	41.64
Iran	0.36	46.38

Table 3. Hit counts and domains for July (N=30).

Country	Hit counts	Domains
United Kingdom	100.00	3.18
Canada	47.74	87.18
France	46.76	81.88
Mainland China	36.35	49.18
Brazil	33.40	92.00
Germany	32.22	4.59
Italy	24.56	94.24
Argentina	15.11	81.06
Mexico	13.67	89.06
Spain	11.00	4.47
South Africa	8.47	57.65
India	7.52	84.82
Turkey	7.43	71.41
United States of America	6.35	3.41
Russia	6.25	81.76
Colombia	4.20	68.00
Peru	4.01	41.18
Sweden	3.61	83.88
Chile	2.85	77.53
Ecuador	1.80	58.59
Indonesia	1.53	100.00
Pakistan	1.07	57.65
Philippines	1.02	61.41
Egypt	0.73	30.59
Iran	0.57	58.47
Bangladesh	0.47	42.47
Kazakhstan	0.26	44.82
Saudi Arabia	0.15	32.94
Qatar	0.05	35.06
Iraq	0.02	25.41

Table 1 shows that the countries with the highest hit counts in February were Australia, Canada, Italy, the United Kingdom, and Germany. Table 2 shows that the countries with the highest hit counts in April were Canada, France, and the United Kingdom. Table 3 shows that the countries with the highest hit counts in July were the United Kingdom, Canada, and France. In February, there were more domains in Australia, Japan, and Singapore. In April, more domains were observed in Sweden, the United States, and the Netherlands. In July, more domains were found in Indonesia, Italy, and Brazil.

Following these results, we analyzed publishing organizations. We visualized the countries involved in the high-frequency second TLD data as a two-mode network. The data revealed 23 countries in February, 26 countries in April, and 26 countries in July. The visualization results are shown in Figures 1-6. The large “G” in the figures denotes “group.” For example, G(GEOC) means the group in which G (government), E (education), O (nonprofit organizations), and C (commercial organizations) appear simultaneously.

As seen in Figure 1, most of the COVID-19–related messages released in February were from the government, educational, or commercial sectors, with relatively few messages from the nonprofit sector. We divided the countries into several groups based on areas. Among all groups, the countries and regions that received information from these four areas the most included mainland China, Hong Kong, Macao, Australia, and Vietnam. Asian countries accounted for 88% (14/16) of these countries. Information from Italy was primarily from the government and educational enterprises. In Sri Lanka and the United Arab Emirates, information was mainly from the

XSL•FO RenderX

government and educational institutes. In the United States and Russia, information was from the government and the educational and commercial sectors. In the United States, government agencies were primarily concentrated in California. Spain reported more commercial agencies, whereas Belgium reported the most information from the educational field. In general, Asian countries were more diverse regarding the online information shared on COVID-19 in February than other regions. This could be because most of the confirmed COVID-19 cases were diagnosed during this period in Asia, and various regions of the continent were considered to be more sensitive than others [42].

Figure 2 shows the institutional network diagram of the COVID-19–related information released in February 2020. The connection between nodes represents the simultaneous release of COVID-19–related information by these institutions. The width of the connection line represents the frequency of coreleases, and the wider the line, the more simultaneous the releases. The bold line between G and C indicates that the government and commercial area released the most information simultaneously, followed by the government and educational sector, and then the commercial and educational sector. Within the framework of the quadruple helix model, the government and the educational and commercial institutions were the leading producers of COVID-19–related information, and they played a prominent role.

Figure 1. Two-mode quadruple helix structure in February 2020. Large "G" refers to the group. G: government domains; E: educational institute domains; C: commercial domains; O: nonprofit organization domains.

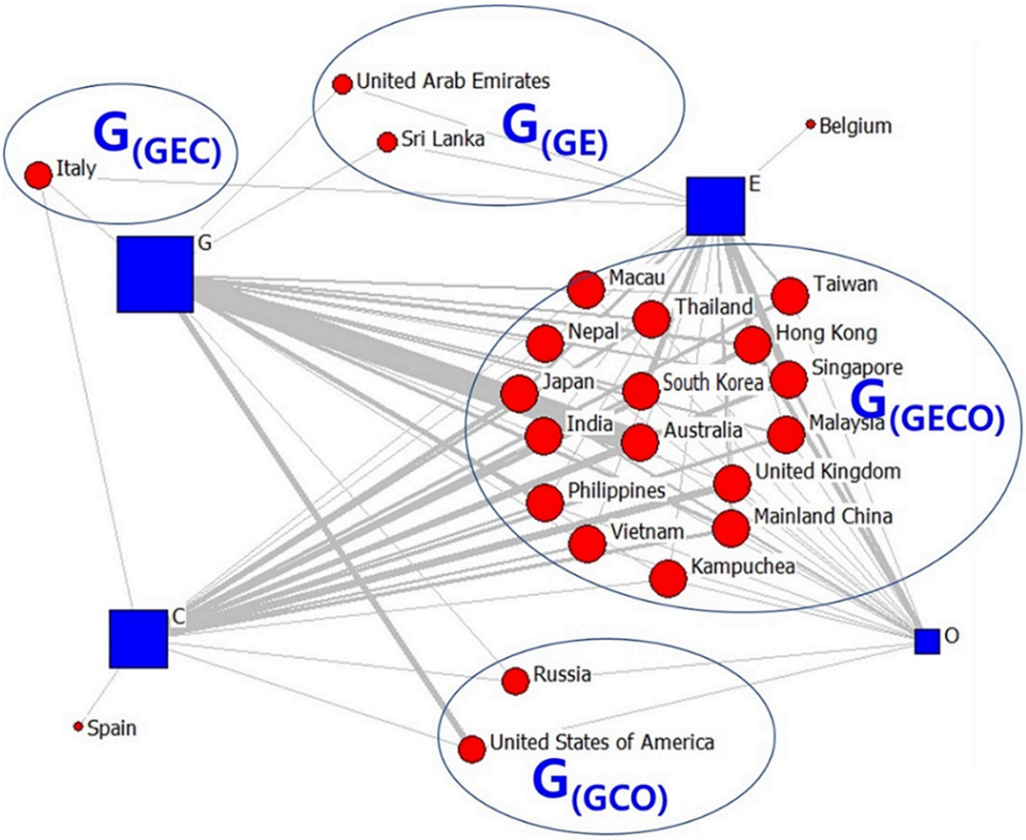
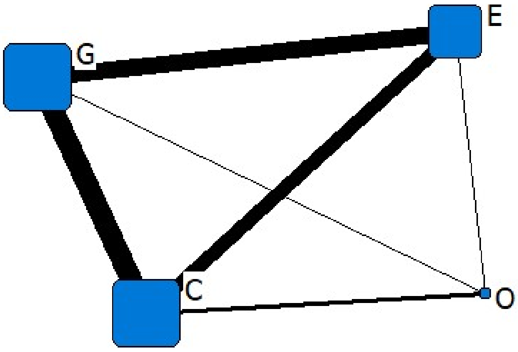

United Arab Emirates
Sri Lanka
G(GE)
Italy
G(GEC)
Belgium
E
G
Macau
Taiwan
Thailand
Nepal
Hong Kong
Singapore
Japan
South Korea
G(GECO)
India
Australia
Malaysia
United Kingdom
Philippines
Mainland China
Vietnam
Kampuchea
C
O
Russia
Spain
United States of America
G(GCO)

Figure 2. One-mode quadruple helix structure in February 2020. G: government domains; E: educational institute domains; C: commercial domains; O: nonprofit organization domains.



As seen in Figure 3, April's COVID-19 messages were primarily from the government and the educational and commercial sectors, and relatively little information was provided by the nonprofit sector, as was the case in February. The number of countries and regions that received information from all four areas simultaneously was the largest. The areas in which COVID-19 information was released in these countries were relatively diverse, including mainland China, the United Kingdom, Brazil, and Japan, with Asian countries accounting for less than half. This is significantly different from the situation in February because, at this stage, COVID-19 became a pandemic and was no longer concentrated in Asia. Information from Romania, Peru, and Chile was primarily from the government and the educational and nonprofit sectors. In Spain, information was mostly from the government and the commercial and nonprofit sectors. Portugal, Italy, India, and Ireland received information primarily from the government and educational sector. Countries where commercial and nonprofit agencies released more information included Israel, the United States, Austria, and Sweden. For South Korea and Poland, government and commercial sectors released more information. In Russia and the Netherlands, most information was shared by government agencies, while in Switzerland and Belgium, educational institutions were the primary sources of COVID-19–related information. In France, the information was primarily shared through the commercial sector. In general, the online information shared about COVID-19 during April and February was quite different in terms of both countries and institutions. Non-Asian countries diversified their fields as COVID-19 became a pandemic.

Figure 4 shows the institutional network diagram of the COVID-19–related information released in April. The government and the educational sector released the most information at this time. The relationship between the government and commercial sector, and that between commercial and nonprofit organizations was also closer. In April, the government and the educational and commercial institutions were still the leading producers of information, playing a prominent role in information dissemination.

Figure 3. Two-mode quadruple helix structure in April 2020. Large "G" refers to the group. G: government domains; E: educational institute domains; C: commercial domains; O: nonprofit organization domains.

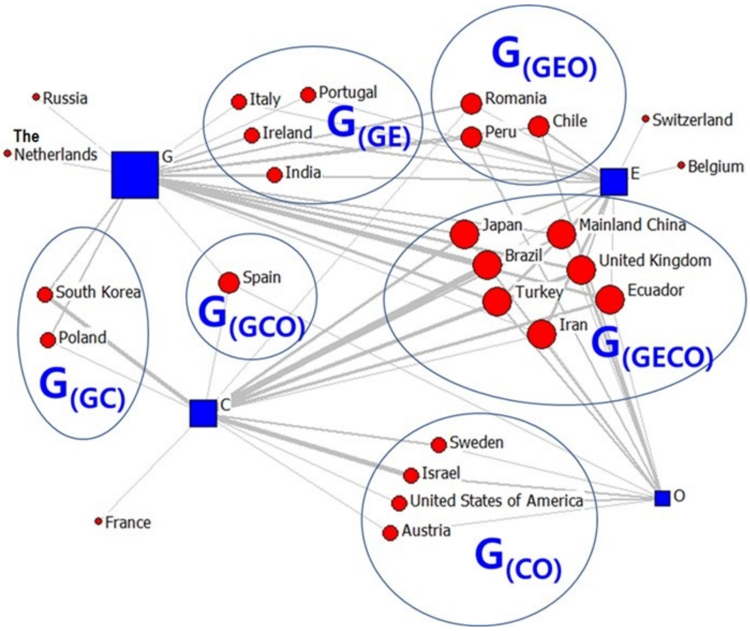
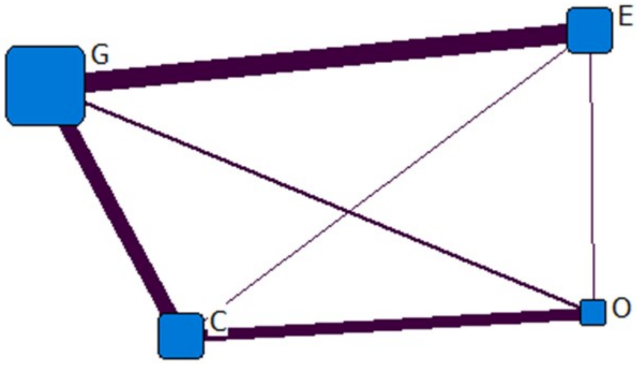


Figure 4. One-mode quadruple helix structure in April 2020. G: government domains; E: educational institute domains; C: commercial domains; O: nonprofit organization domains.



The two-mode diagram of the COVID-19 information release in July shows that most countries were delivering information from a diverse range of sectors (Figure 5). The proportion of individual areas and of countries and regions that received information from only two areas was lower than that in the previous phases. The countries and regions that received information from all four sectors the most included mainland China, Russia, Turkey, and the Philippines. In April, the COVID-19 pandemic continued to spread around the world, and the geographic distribution of information was also seen globally, and not just in Asia. Information from Iraq was primarily from the government and commercial and educational organizations. Information in France was from the government and nonprofit sectors. Government agencies in Chile and Italy provided relatively more information. In the United States, information from the government and educational sector decreased, while information from commercial sectors increased. Less information was collected from Bing in the United States in July. Most regions of the world and many industries were affected by the pandemic in July. The structure of the network for information-publishing organizations also developed from the coexistence of double, triple, and quadruple helices to the main structure of quadruple helices.

Figure 6 shows the institutional network diagram representing the COVID-19–related information released in July. The number of concurrent announcements made by the government and commercial sector remained the highest, followed by the government and educational sector, and then the commercial and educational sectors. In the three stages, the government and the educational and commercial institutions were the leading producers of information and played a prominent role.

We collated the degree centralities in four helices and found that the commercial sector in February had the highest degree, followed by the government and educational sector, and finally the nonprofit organizations (Figure 7). In April, the biggest area of degree centrality was again the commercial sector, followed by the government, nonprofit organizations, and finally educational organizations. In July, the government ranked first, commercial organizations ranked second, educational organizations were third, and the nonprofit sector fourth. Thus, the government and commercial organizations played a significant role in the COVID-19 information network, whereas the role of the nonprofit sector was relatively small.

Figure 5. Two-mode quadruple helix structure in July 2020. Large "G" refers to the group. G: government domains; E: educational institute domains; C: commercial domains; O: nonprofit organization domains.

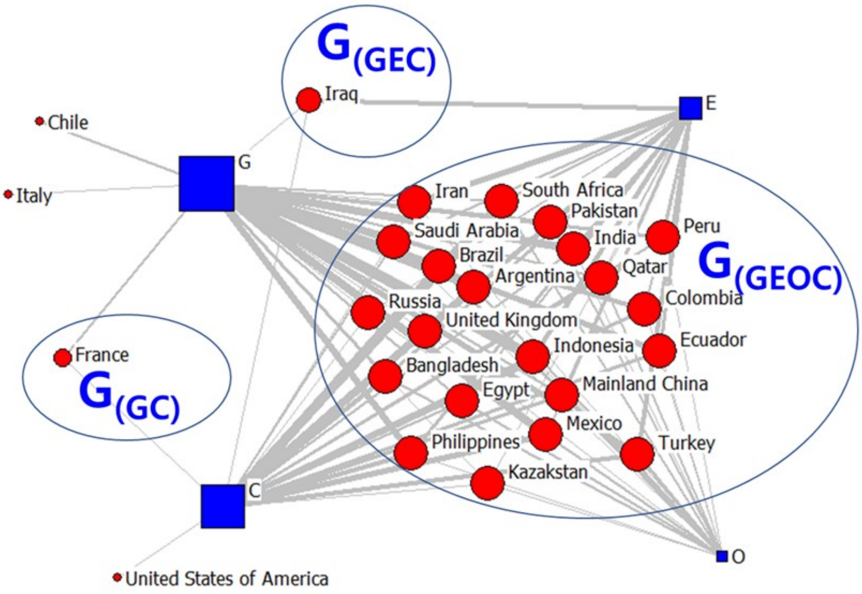


Figure 6. One-mode quadruple helix structure in July 2020. G: government domains; E: educational institute domains; C: commercial domains; O: nonprofit organization domains.

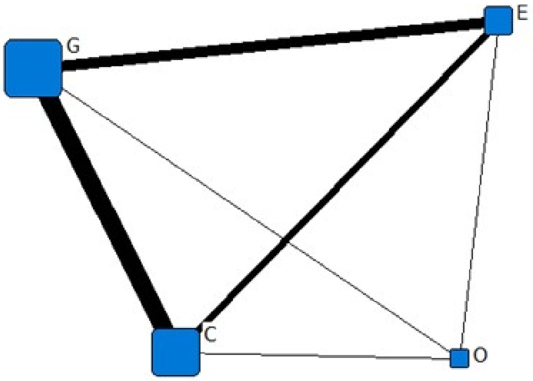
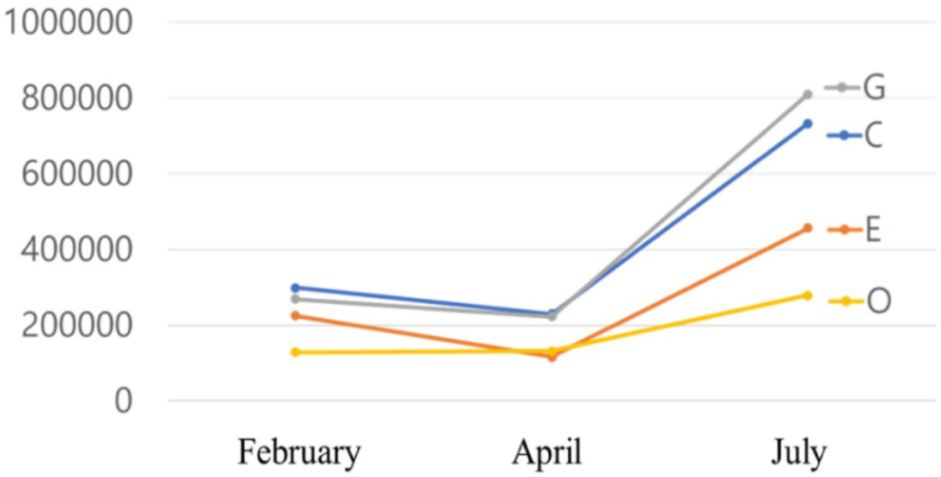


Figure 7. Degrees of the three stages. G: government domains; E: educational institute domains; C: commercial domains; O: nonprofit organization domains.



We performed a text analysis and CONCOR analysis for the content of the information shared. For the content analysis, we deleted non-English and scrambled characters during data cleaning. There was a total of 9149 documents in February, and 8889 remained after cleaning. There were originally 14,768 documents in April, 14,766 of which remained after cleaning. The number of documents in July was 14,484 and 13,087 remained after cleaning. Word preprocessing was first performed using Python (the Spacy package) and the results were manually collated. We identified the 50 most frequently found words during each of the three months, which are compiled in Table 4.

Table 4 indicates that in February, people paid the most attention to the affected areas (China), health, departments, international, and news. In April, the content was focused on information regarding deaths, health, the pandemic, and public. In July, the content was focused on information regarding the pandemic, health, and news. These remained the top concerns in July 2020, whereas words such as “online,” “service,” “university,” and “government” were also highly ranked at this time. To summarize, the main content in February was dominated by information and news about the outbreak; in April, information was primarily regarding the public and the pandemic; and in July, various online services were used to address the problems caused by the pandemic.

Table 4. The 50 most frequent words of the three stages.

Rank	February		April		July	
	Words	Standardized frequency	Words	Standardized frequency	Words	Standardized frequency
1	coronavirus	100.00	coronavirus	100.00	coronavirus	100.00
2	country	59.30	information	24.96	information	36.79
3	China	30.18	die	21.70	pandemic	24.33
4	novel	28.57	health	12.09	health	22.67
5	health	27.90	virus	9.96	online	18.62
6	Jan	21.14	pandemic	8.22	service	17.20
7	international	19.88	public	7.64	virus	14.21
8	year	19.53	update	6.88	university	12.67
9	department	17.85	country	6.69	government	12.24
10	news	16.35	pour	6.62	news	12.12
11	world	16.00	school	6.60	provide	11.88
12	school	15.91	service	6.58	case	11.28
13	new	15.24	meet	5.91	July	10.49
14	visit	14.22	spread	5.69	new	9.98
15	statement	13.62	April	5.53	development	9.73
16	spread	13.30	government	5.24	education	9.46
17	patient	12.65	March	5.16	ministry	9.46
18	NSW	12.44	SARS	4.92	update	9.46
19	case	10.78	China	4.91	June	9.43
20	epidemic	10.37	situation	4.86	world	9.22
21	response	9.04	help	4.52	country	8.98
22	Japan	8.92	case	4.51	China	8.49
23	student	8.92	work	4.51	support	8.49
24	find	8.81	student	4.48	work	8.31
25	official	8.81	new	4.47	help	8.28
26	category	8.72	continue	4.36	business	8.19
27	university	8.52	provide	4.36	time	7.83
28	staff	8.40	novel	4.13	public	7.59
29	information	7.98	outbreak	4.13	spread	7.35
30	symptom	7.87	community	4.01	March	7.29
31	child	7.43	care	3.91	student	7.29
32	like	6.96	support	3.91	include	7.26
33	continue	6.82	online	3.88	Pakistan	7.10
34	city	6.71	people	3.75	SARS	7.04
35	public	6.67	website	3.75	India	6.86
36	Feb	6.59	maatregelen	3.67	Indonesia	6.77
37	ministry	6.12	include	3.64	community	6.71
38	disease	5.98	medidas	3.53	website	6.56
39	big	5.97	time	3.52	year	6.50
40	monitor	5.95	business	3.38	measure	6.47
41	pneumonia	5.83	impact	3.23	continue	6.44

Rank	February		April		July	
	Words	Standardized frequency	Words	Standardized frequency	Words	Standardized frequency
42	update	5.75	man	3.11	disease	6.41
43	government	5.36	find	3.08	national	6.41
44	service	5.33	university	3.07	people	6.35
45	unfold	5.07	page	3.06	find	6.32
46	view	4.81	contact	3.02	medidas	6.32
47	Singapore	4.63	disease	2.98	home	6.26
48	cause	4.60	staff	2.94	social	6.14
49	infect	4.58	late	2.93	South	6.14
50	Chinese	4.29	measure	2.92	read	6.08

To further assimilate valuable information, we constructed the semantic network of high-frequency words and used CONCOR analysis for clustering. Furthermore, we obtained a visualization diagram of the clustering network in the three stages.

Group 1 in Figure 8 is named "Ministry and University," which contains information on services provided by the government departments, public authorities, and schools. Group 2 is named "Virus Spreading," which primarily includes confirmed cities, patient symptoms, and the spread of the infection. Group 3 is "Coronavirus and Health" and Group 4 is "Child and Education." Most of the information in Group 4 is related to "NSW (New South Wales, Australia) education" and "child," and it showed the highest frequency in February.

Figure 8. Semantic network in February 2020.

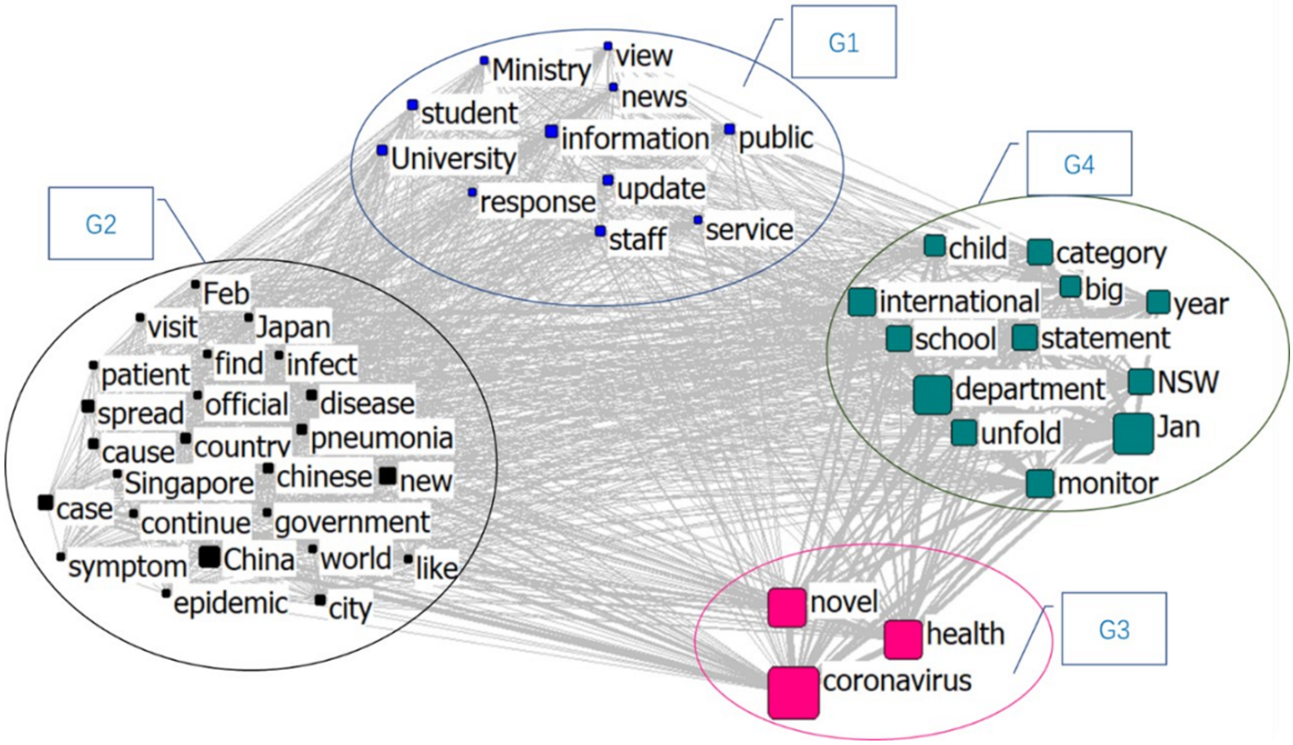
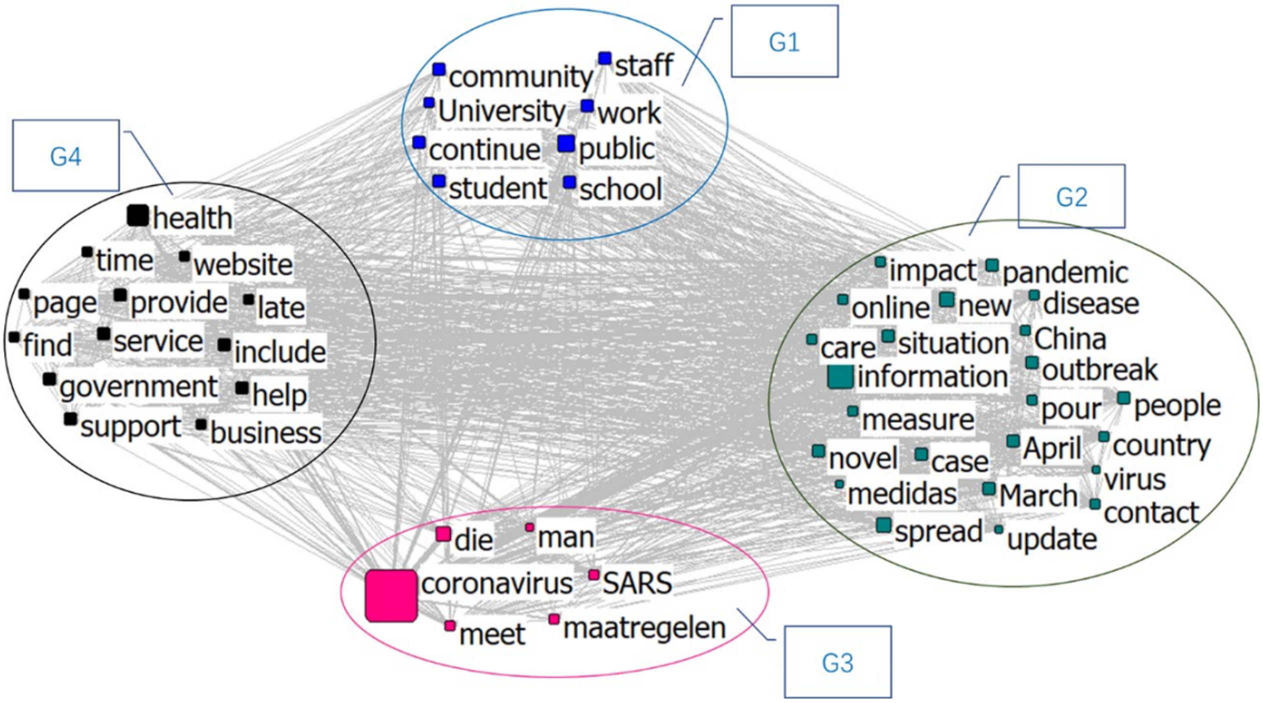

G1
Ministry
view
news
student
information
public
University
update
response
staff
service
G2
Feb
visit
Japan
patient
find
infect
official
disease
spread
cause
country
pneumonia
Singapore
chinese
new
case
continue
government
symptom
China
world
like
epidemic
city
G4
child
category
international
big
year
school
statement
department
NSW
unfold
Jan
monitor
novel
health
coronavirus
G3

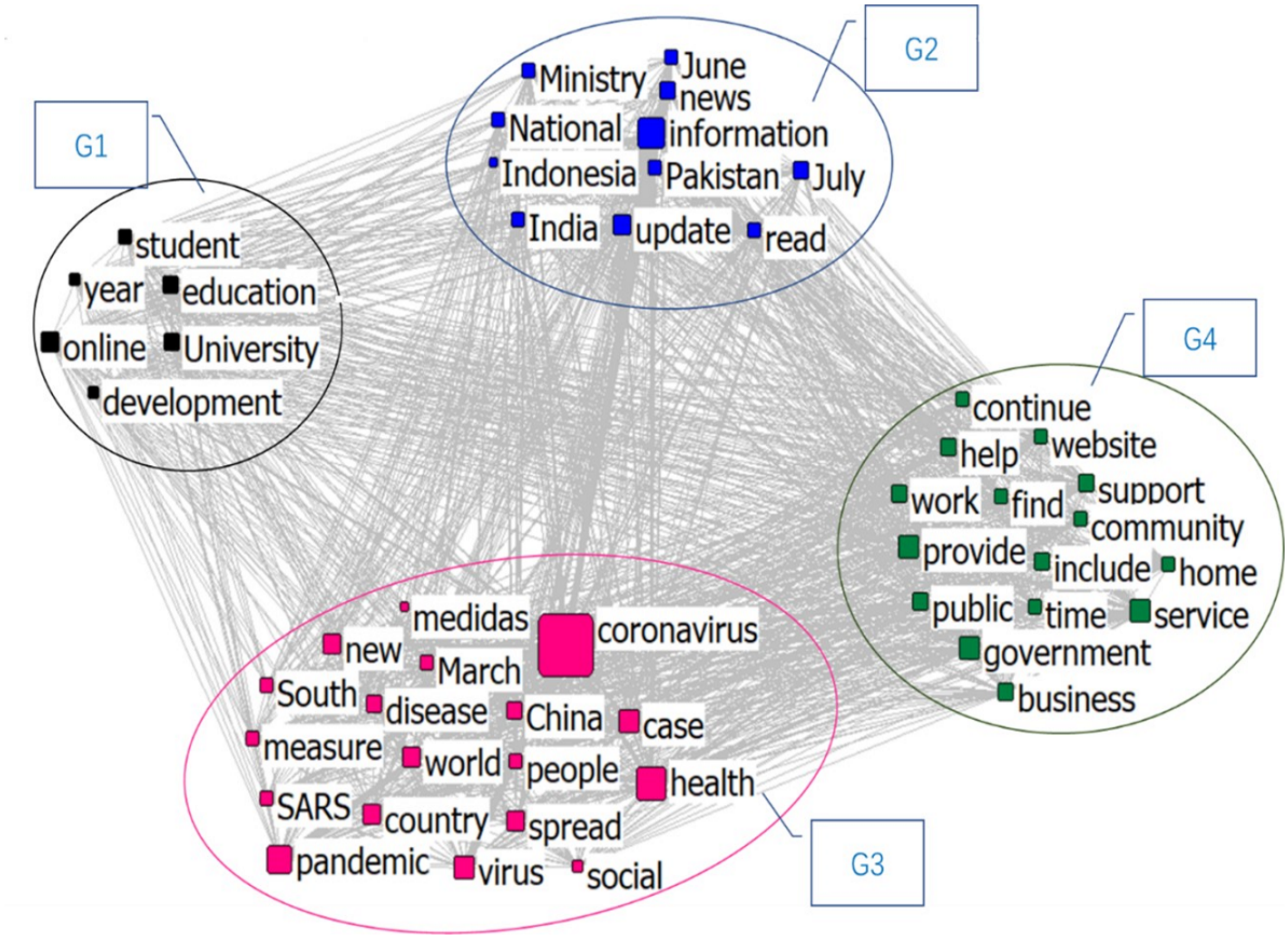
Figure 9 shows the semantic network for April. Group 1 is a school and student-related group named "Education Issue." Group 2 is "Virus Spreading," including information about the outbreak and the spread of the pandemic. Group 3 is "Virus Description," which contains information related to the characteristics of the virus. Group 4 is "Commercial Issue," which includes words such as "business," "government," "service," and "help," and is related to the social change brought about by the pandemic.

Figure 9. Semantic network in April 2020.


G1
community
staff
University
work
continue
public
student
school
G4
health
time
website
page
provide
late
find
service
include
government
help
support
business
G2
impact
pandemic
online
new
disease
care
situation
China
information
outbreak
measure
pour
people
novel
case
April
country
medidas
March
virus
contact
spread
update
die
man
coronavirus
SARS
meet
maatregelen
G3

In July, the semantic network was also divided into four groups, as shown in Figure 10. Group 1 is "Distance Education," which contains information about online education. Group 2 is "City News," which contains information about the cities affected by the pandemic. Group 3 contains information about the measures taken, and is therefore named "Measures." The last group is "Commercial Issues," which includes information regarding "business," "services," "government," "provide," "community," and similar.

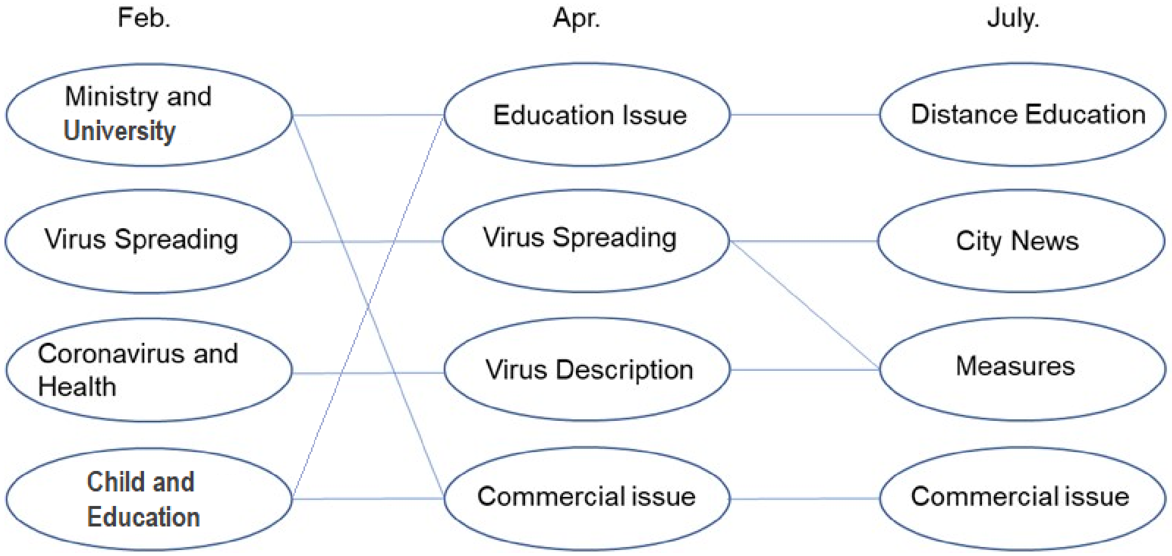
Figure 10. Semantic network in July 2020.


G2
Ministry
June
news
National
information
Indonesia
Pakistan
July
India
update
read
G1
student
year
education
online
University
development
G4
continue
help
website
work
find
support
community
provide
include
home
public
time
service
government
business
medidas
coronavirus
new
March
South
disease
China
case
measure
world
people
health
SARS
country
spread
pandemic
virus
social
G3

The clustering topics of the three stages are sorted in Figure 11. The results show that the information in February was mainly regarding the response of the government and educational institutes, such as the impact on schools after the virus began spreading, and the spread of the virus and health issues became prominent topics. In addition, since the spread of COVID-19 was at an early stage in February, information reports in some places were also relatively prominent. By April, the disease had become a pandemic. At this stage, besides information, the spread of the virus, the description of the related characteristics of the virus, and the changes of the commercial environment became prominent topics. By July, the focus in education shifted to distance education. As the pandemic could not be fully controlled within a short period of time, most educational institutions began to prepare for or implement online education. By this time, the public had a basic understanding of the virus and how it spreads, and the focus shifted to measures such as how to deal with this spread. The impact of the pandemic on business and society was still an important topic. Information about cities related to the outbreak also continued to appear in the news.

In general, education became a prominent topic of discussion in all three stages. With time, the basic information regarding the virus and its transmission became popularized, and people began to pay more attention to information about measures to prevent its spread. Since the beginning of the pandemic, the situation has changed in terms of business, government policy, and other public issues. Society has also changed. We compared the results of content analysis with the results of the quadruple helix structure and found that the content analysis also confirmed the form of the quadruple helix structure. In the content analysis, the information groups about business issues and government emerged as relatively large, with a smaller contribution of information about education, although this topic also forms a certain scale of the groups.

Figure 11. Evolution of topic content.



Discussion

Principal Findings

In this study, we analyzed COVID-19 web information sources from a quadruple helix perspective, and found changes in structure and content at each stage during the first 6 months of online information regarding COVID-19. We also found problems in the structure of information sources in the transmission of relevant information. We here provide detailed suggestions, which can contribute to the internet communication of future public health events.

Based on the quadruple helix model, this study collated and analyzed the structure and content of the network information sources about COVID-19 considering time and space. By sorting out the second TLD, we divided the structure of network information sources into four categories: the government, education, companies, and nonprofit organizations. An information source network composed of four levels was obtained. The results of the two-mode quadruple helix analysis of the three stages showed that the major confirmed cases in February (first stage) occurred in Asia, and the online information sources in Asia were more diversified than those in other regions. As the pandemic spread in April (second stage), non-Asian sources of information began to diversify, and in July (third stage), the sources of web information became globalized. Thus, the impact of the first stage of the pandemic was more sensitive in Asia, and the information from various industries was related to responding to this need. However, only some industries in non-Asian regions paid attention in the first stage, and the information source helix did not form, which also led to the slow response to COVID-19 in some regions and the delay in response measures [43]. Since April, the spiral has intensified in non-Asian regions due to the spread of the pandemic to many areas outside Asia, which has raised concern of various industries.

In general, from the results of the two-mode analysis, the structure of the three stages of web information publishing organization has gradually developed from the coexisting structure of a double, triple, and quadruple helix to the diversified structure centered on the triple and quadruple helix. From this phenomenon, we can find that in the face of major public health emergencies, most of the local information release sources are not comprehensive. This phenomenon has also led to the failure of many industries to anticipate and respond to the pandemic in a timely fashion [44,45]. Our results suggest that the health care sector can call on the local information sources of various industries to release appropriate and reasonable information about the health and public events in the future to ensure the timely deployment of all sectors of society and avoid more losses. We used a modular quadruple helix structure to analyze the forces of these four levels at various periods in detail. We found that in February, the information shared was the most coincident and closely linked between government and commercial organizations, followed by educational and government organizations. Next, there was a closeness between the commercial and educational sectors. In April, the government and the educational sector simultaneously released the most information about COVID-19. The relationship between the government and commercial sectors, and the relationship between commercial and nonprofit sectors were also closer. In July, the number of concurrent announcements about COVID-19 by the government and commercial sector remained the highest, followed by the government and educational sector, and then the commercial and educational sectors. We collated the centrality of the three stages and four areas, indicating that the commercial area scored the highest in February, followed by the government and educational sectors, and finally the nonprofit organizations. In April, the biggest area of degree centrality was also the commercial sector, followed by the government, nonprofit sector, and finally educational enterprises. In July, during the third stage, the government played a central role in the COVID-19 information network. In all three stages, as a whole, the government and commercial sector played a significant role in COVID-19 network information, and the connection of the nonprofit sector was relatively low. In fact, in the event of major infectious diseases, school is an important aspect that cannot be ignored, and schools often gather dense populations [46]. The communication role of the education sector as an information source is not stronger than that of business and government sectors. However, as educational institutions learn more than any other institutions about the actual school and education situation, they should take on more of a role than the government and businesses to ensure the spread of information. In future infectious disease health events, education and industry organizations, along with others, need to release information more quickly and accurately.

This study included an analysis of the quadruple helix structure and the content of the three stages using dynamic progressive detailed analysis. We carried out content analysis on 36,742 pieces of information in the three stages. The results of frequency analysis showed that the most prominent information in February was news about the pandemic. April was dominated by information about the public and the pandemic. The focus in July was the use of various online services to solve problems caused by the pandemic. We then used CONCOR cluster analysis to classify the topics in the three stages. The changes in trends in the three stages were also sorted. The results indicated that in the early stages, there were more reports about the affected areas, and the response of authorities such as governments and schools to the virus, and the spread of the virus and health issues were the main points of focus of discussion. The second phase focused on the spread of the virus across the world, which created a global pandemic. At this stage, information about educational hotspots, descriptions of virus-related features, and information about commercial environment changes caused by the pandemic also received attention. In the third stage, the educational hotspots differentiated into the characteristics of distance education. The pandemic made physical face-to-face education difficult. Many educational institutions began to prepare for or implement online education. Public attention at this stage shifted from what the virus was to measures of controlling its spread. In general, education was a prominent topic at all levels. With the change of stage, the information content also changed. In the early stage, the basic situation of the virus and its impact on health attracted most of the attention. Later, the focus was on pandemic prevention measures. The business environment and policy environment have changed from the beginning of the pandemic, and the social changes caused by the pandemic have also become an important discussion topic.

Limitations

Owing to the large amount of data from all countries worldwide, this study has only used the web information for countries with a significant number of diagnosed cases at each stage as the research object. In addition, we only used data from Bing. Although Bing is more widely used than any other search engine in the webometric field, it does not have a strong market share in some parts of the world that rely more on other search engines. For example, Google has the largest market share in the United States, Baidu has the largest share in China, and Naver has the largest share in South Korea. Therefore, the results of different search engines in individual regions may somewhat vary from those of Bing. In addition, there is no ideal description of the web network structure [47]. Search engine properties are considered as more engineering products than mathematical tools [48]. Different search engines often have divergent algorithms and search results, which inevitably produce repeated and mixed results. Since search engines usually consider both quality and efficiency, this could also lead to problems related to Type I and Type II errors, which objectively lead to insufficient coverage [48,49]. These can be considered as limitations of the study.

Conclusions

This study focused on the structure of information sources at each stage of the first 6 months of the COVID-19 pandemic and the development of the network structure through the quadruple helix framework. We found that for public health emergencies, some online and offline information sources were not sufficient. Diversified institutions need to pay attention to public health emergencies, and actively respond to multihelical information sources, which is conducive to implementing a timely and more comprehensive response to public health emergencies. In terms of published messages, the educational sector plays an important role in public health events. However, educational institutions release less information than governments and businesses. In addition, we summarized the trend of COVID-19 online information dissemination. It is important to understand the communicational structure of pandemic information sources worldwide. Currently, the quadruple helix model is primarily used in the field of scientific cooperation in terms of coauthorship analysis, and research in other fields is insufficient. This study highlights that the quadruple helix not only has theoretical significance in the scientific innovation field but can be also used to conduct effective research regarding web information. This is significant for further development of the quadruple helix model with respect to the COVID-19 pandemic.

Acknowledgments

This research is partially funded by the Foundation for Broadcast Culture (방송문화진흥회) in the Republic of Korea. The data for February were analyzed and reported from a different scheme for a Korean-language journal. Since then, the data for April and July have been compiled only for this study, as COVID-19 has rapidly become more serious over time.

Conflicts of Interest

None declared.

References

1. COVID-19 Dashboard. Center for Systems Science and Engineering (CSSE), Johns Hopkins University. URL: https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299%0A423467b48e9ecf6%0D [accessed 2021-01-07]
2. Kim J, Ashihara K. National disaster management system: COVID-19 case in Korea. Int J Environ Res Public Health 2020 Sep 14;17(18):e6691 [FREE Full text] [doi: 10.3390/ijerph17186691] [Medline: 32937946]
3. Horrigan JB. Information overload. Pew Research Center. 2016 Dec 07. URL: https://www.pewresearch.org/internet/2016/12/07/information-overload [accessed 2021-01-06]
4. Park HW, Kim JE, Zhu YP. Online information sources of coronavirus using webometric big data. J Korea Acad Ind Coop Soc 2020 Nov 30;21(11):728-739 In Korean. [doi: 10.5762/KAIS.2020.21.11.728]
5. Kink N, Hess T. Search engines as substitutes for traditional information sources? An investigation of media choice. Inf Society 2008 Jan 14;24(1):18-29. [doi: 10.1080/01972240701771630]
6. Kitamura S. The relationship between use of the internet and traditional information sources. SAGE Open 2013 May 21;3(2):215824401348969. [doi: 10.1177/2158244013489690]
7. Han J, Cha M, Lee W. Anger contributes to the spread of COVID-19 misinformation. HKS Misinfo Rev 2020 Sep 17;1:1-14. [doi: 10.37016/mr-2020-39]
8. Calvillo DP, Ross BJ, Garcia RJB, Smelter TJ, Rutchick AM. Political ideology predicts perceptions of the threat of COVID-19 (and susceptibility to fake news about it). Soc Psychol Person Sci 2020 Jul 22;11(8):1119-1128. [doi: 10.1177/1948550620940539]
9. Marco-Franco JE, Pita-Barros P, Vivas-Orts D, González-de-Julián S, Vivas-Consuelo D. COVID-19, fake news, and vaccines: should regulation be implemented? Int J Environ Res Public Health 2021 Jan 16;18(2):e744 [FREE Full text] [doi: 10.3390/ijerph18020744] [Medline: 33467179]
10. Casero-Ripolles A. Impact of Covid-19 on the media system. Communicative and democratic consequences of news consumption during the outbreak. EPI 2020 Apr 23;29(2):e290223. [doi: 10.3145/epi.2020.mar.23]
11. Park HW, Chung SW. Editor's note response to Friedman's "The world before corona and the world after": A perspective raging from the development of civilization to the harmony of east and west, and the paradigm shift. J Contemp East Asia 2020 Dec 31;19(2):169-178. [doi: 10.17477/jcea.2020.19.2.169]
12. Chong M, Park HW. COVID-19 in the Twitterverse, from epidemic to pandemic: information-sharing behavior and Twitter as an information carrier. Scientometrics 2021 Jun 23;126:1-25 [FREE Full text] [doi: 10.1007/s11192-021-04054-2] [Medline: 34188332]
13. Kuchler T, Russel D, Stroebel J. JUE Insight: The geographic spread of COVID-19 correlates with the structure of social networks as measured by Facebook. J Urban Econ 2021 Jan 9:103314. [doi: 10.1016/j.jue.2020.103314]
14. Li D, Ko N, Chen Y, Wang P, Chang Y, Yen C, et al. COVID-19-related factors associated with sleep disturbance and suicidal thoughts among the Taiwanese public: a Facebook survey. Int J Environ Res Public Health 2020 Jun 22;17(12):e4479 [FREE Full text] [doi: 10.3390/ijerph17124479] [Medline: 32580433]
15. Xue J, Chen J, Hu R, Chen C, Zheng C, Su Y, et al. Twitter discussions and emotions about the COVID-19 pandemic: machine learning approach. J Med Internet Res 2020 Nov 25;22(11):e20550 [FREE Full text] [doi: 10.2196/20550] [Medline: 33119535]
16. Cho SE, Park HW. Government organizations’ innovative use of the Internet: The case of the Twitter activity of South Korea’s Ministry for Food, Agriculture, Forestry and Fisheries. Scientometrics 2011 Sep 28;90(1):9-23. [doi: 10.1007/s11192-011-0519-2]

17. Thelwall M. Does astronomy research become too dated for the public? Wikipedia citations to astronomy and astrophysics journal articles 1996-2014. EPI 2016 Nov 14;25(6):893-900. [doi: 10.3145/epi.2016.nov.06]
18. Park SJ, Lim YS, Park HW. Comparing Twitter and YouTube networks in information diffusion: The case of the “Occupy Wall Street” movement. Technol Forecast Soc Change 2015 Jun;95:208-217. [doi: 10.1016/j.techfore.2015.02.003]
19. Park HW, Lim YS. Do North Korean social media show signs of change?: An examination of a YouTube channel using qualitative tagging and social network analysis. J Contemp East Asia 2020 Jul 31;19(1):123-143. [doi: 10.17477/jcea.2020.19.1.123]
20. Jung K, Park HW. Tracing interorganizational information networks during emergency response period: A webometric approach to the 2012 Gumi chemical spill in South Korea. Gov Inf Quart 2016 Jan;33(1):133-141. [doi: 10.1016/j.giq.2015.09.010]
21. Allaire MC. Disaster loss and social media: can online information increase flood resilience? Water Resour Res 2016 Sep 29;52(9):7408-7423. [doi: 10.1002/2016wr019243]
22. Park HW. YouTubers’ networking activities during the 2016 South Korea earthquake. Qual Quant 2017 Mar 27;52(3):1057-1068. [doi: 10.1007/s11135-017-0503-x]
23. Kim J, Bae J, Hastak M. Emergency information diffusion on online social media during storm Cindy in U.S. Int J Inf Manag 2018 Jun;40:153-165. [doi: 10.1016/j.ijinfomgt.2018.02.003]
24. Song J, Song TM, Seo D, Jin D, Kim JS. Social big data analysis of information spread and perceived infection risk during the 2015 Middle East Respiratory Syndrome outbreak in South Korea. Cyberpsychol Behav Soc Netw 2017 Jan;20(1):22-29. [doi: 10.1089/cyber.2016.0126] [Medline: 28051336]
25. 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]
26. O'Neil M, Raissi M, Turner B. The case for asymmetry in online research: Caring about issues in Australian and Canadian Web 1.0 bee networks. Int J Commun 2020;14:5150-5173 [FREE Full text]
27. Lu Y, Chen Y. Hierarchical cluster and multidimensional scaling analysis of video websites based on URL co-occurrence. 2020 Sep 22 Presented at: 2020 IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS); July 28-30, 2020; Shenyang, China p. 590-594. [doi: 10.1109/icpics50287.2020.9201979]
28. Bianchi G, Bruni R, Daraio C, Laureti Palma A, Perani G, Scalfati F. Exploring the potentialities of automatic extraction of university webometric information. J Data Inf Sci 2020 Nov 21;5(4):43-55. [doi: 10.2478/jdis-2020-0040]
29. Park S, Park HW. Diffusion of cryptocurrencies: web traffic and social network attributes as indicators of cryptocurrency performance. Qual Quant 2019 Jan 25;54(1):297-314. [doi: 10.1007/s11135-019-00840-6]
30. Etzkowitz H, Leydesdorff L. The triple helix -- university-industry-government relations: A laboratory for knowledge based economic development. EASST Rev 1995 Jan 1;14(1):14-19 [FREE Full text]
31. Leydesdorff L. The triple helix, quadruple helix, …, and an n-tuple of helices: explanatory models for analyzing the knowledge-based economy? J Knowl Econ 2011 Jun 18;3(1):25-35. [doi: 10.1007/s13132-011-0049-4]
32. Carayannis EG, Campbell DFJ. 'Mode 3' and 'Quadruple Helix': toward a 21st century fractal innovation ecosystem. Int J Technol Manag 2009 Feb 20;46(3/4):201-234. [doi: 10.1504/ijtm.2009.023374]
33. Carayannis EG, Campbell DFJ. Triple helix, quadruple helix and quintuple helix and how do knowledge, innovation and the environment relate to each other?: A proposed framework for a Ttrans-Disciplinary analysis of sustainable development and social ecology. Int J Soc Ecol Sustain Dev 2010;1(1):41-69. [doi: 10.4018/jsesd.2010010105]
34. Zhu Y, Park HW. Uncovering blockchain research publications in Asia compared to the rest of the world. Korean Data Anal Soc 2020 Apr 30;22(2):513-526. [doi: 10.37727/jkdas.2020.22.2.513]
35. Park HW, Yoon J. Structural characteristics of institutional collaboration in North Korea analyzed through domestic publications. Scientometrics 2019 Mar 9;119(2):771-787. [doi: 10.1007/s11192-019-03056-5]
36. Park S, Chung D, Park HW. Analytical framework for evaluating digital diplomacy using network analysis and topic modeling: Comparing South Korea and Japan. Inf Process Manag 2019 Jul;56(4):1468-1483. [doi: 10.1016/j.ipm.2018.10.021]
37. Park HW, Leydesdorff L. Decomposing social and semantic networks in emerging “big data” research. J Informetrics 2013 Jul;7(3):756-765. [doi: 10.1016/j.joi.2013.05.004]
38. Yoon J, Yang JS, Park HW. Quintuple helix structure of Sino-Korean research collaboration in science. Scientometrics 2017 Jul 27;113(1):61-81. [doi: 10.1007/s11192-017-2476-x]
39. Borgatti SP, Everett MG. A graph-theoretic perspective on centrality. Soc Netw 2006 Oct;28(4):466-484. [doi: 10.1016/j.socnet.2005.11.005]
40. Freeman LC. Centrality in social networks conceptual clarification. Soc Netw 1978 Jan;1(3):215-239. [doi: 10.1016/0378-8733(78)90021-7]
41. Chung CJ, Biddix JP, Park HW. Using digital technology to address confirmability and scalability in thematic analysis of participant-provided data. Qual Rep 2020 Sep 11;25(9):3298-3311 [FREE Full text] [doi: 10.46743/2160-3715/2020.4046]
42. Danowski JA, Park HW. East Asian communication technology use and cultural values. J Contemp East Asia 2020 Jul 31;19(1):43-58. [doi: 10.17477/JCEA.2020.19.1.043]

43. Toshkov D, Carroll B, Yesilkagit K. Government capacity, societal trust or party preferences: what accounts for the variety of national policy responses to the COVID-19 pandemic in Europe? J Europ Pub Policy 2021 May 17:1-20. [doi: 10.1080/13501763.2021.1928270]
44. Hanage WP, Testa C, Chen JT, Davis L, Pechter E, Seminario P, et al. COVID-19: US federal accountability for entry, spread, and inequities-lessons for the future. Eur J Epidemiol 2020 Nov;35(11):995-1006 [FREE Full text] [doi: 10.1007/s10654-020-00689-2] [Medline: 33136249]
45. Antwi SH, Getty D, Linnane S, Rolston A. COVID-19 water sector responses in Europe: A scoping review of preliminary governmental interventions. Sci Total Environ 2021 Mar 25;762:143068 [FREE Full text] [doi: 10.1016/j.scitotenv.2020.143068] [Medline: 33131871]
46. Gemmetto V, Barrat A, Cattuto C. Mitigation of infectious disease at school: targeted class closure vs school closure. BMC Infect Dis 2014 Dec 31;14:695 [FREE Full text] [doi: 10.1186/s12879-014-0695-9] [Medline: 25595123]
47. Barnett GA, Park HW. Examining the international internet using multiple measures: new methods for measuring the communication base of globalized cyberspace. Qual Quant 2012 Sep 29;48(1):563-575. [doi: 10.1007/s11135-012-9787-z]
48. Thelwall M. Extracting accurate and complete results from search engines: Case study windows live. J Am Soc Inf Sci 2007 Jan 01;59(1):38-50. [doi: 10.1002/asi.20704]
49. Park HW. How do social scientists use link data from search engines to understand internet-based political and electoral communication? Qual Quant 2011 Jan 1;46(2):679-693. [doi: 10.1007/s11135-010-9421-x]

Abbreviations

CONCOR: convergence of iterated correlations
TLD: top-level domain

Edited by C Basch; submitted 02.02.21; peer-reviewed by Q Liu, R Sauvayre; comments to author 30.03.21; revised version received 16.06.21; accepted 10.07.21; published 26.08.21.

Please cite as:
Zhu YP, Park HW
Development of a COVID-19 Web Information Transmission Structure Based on a Quadruple Helix Model: Webometric Network Approach Using Bing
J Med Internet Res 2021;23(8):e27681
URL: https://www.jmir.org/2021/8/e27681
doi:10.2196/27681
PMID:34280119

©Yu Peng Zhu, Han Woo Park. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Online Public Attention Toward Premature Ejaculation in Mainland China: Infodemiology Study Using the Baidu Index

Shanzun Wei[1], MD; Ming Ma[2], MD; Xi Wen[3], BA; Changjing Wu[3], MA; Guonian Zhu[4], MA; Xiangfu Zhou[1], MD

[1]Department of Urology, The Third Affiliated Hospital, Sun Yat-Sen University, Guangzhou, China

[2]Department of Urology, West China Hospital, Sichuan University, Chengdu, China

[3]Andrology Laboratory, West China Hospital, Sichuan University, Chengdu, China

[4]Institutes for Systems Genetics, Frontiers Science Center for Disease-Related Molecular Network, West China Hospital, Sichuan University, Chengdu, China

Corresponding Author:
Xiangfu Zhou, MD
Department of Urology
The Third Affiliated Hospital
Sun Yat-Sen University
600 Tianhe Rd
Guangzhou, 510630
China
Phone: 86 18227624438
Email: 182119319@QQ.com

Abstract

Background: Premature ejaculation (PE) is one of the most described psychosocial stress and sexual complaints worldwide. Previous investigations have focused predominantly on the prospective identification of cases that meet researchers' specific criteria. The genuine demand from patients with regard to information on PE and related issues may thus be neglected.

Objective: This study aims to examine the online search trend and user demand related to PE on a national and regional scale using the dominant major search engine in mainland China.

Methods: The Baidu Index was queried using the PE-related terms for the period of January 2011 to December 2020. The search volume for each term was recorded to analyze the search trend and demographic distributions. For user interest, the demand and trend data were collected and analyzed.

Results: Of the 36 available PE search keywords, 4 PE searching topics were identified. The Baidu Search Index for each PE topic varied from 46.30% (86,840,487/187,558,154) to 6.40% (12,009,307/187,558,154). The annual percent change (APC) for the complaint topic was 48.80% ($P<.001$) for 2011 to 2014 and –16.82% ($P<.001$) for 2014 to 2020. The APC for the inquiry topic was 16.21% ($P=.41$) for 2011 to 2014 and –11.00% ($P<.001$) for 2014 to 2020. For the prognosis topic, the annual APC was 11.18% ($P<.001$) for 2011 to 2017 and –19.86% ($P<.001$) for 2017 to 2020. For the treatment topic, the annual APC was 14.04% ($P<.001$) for 2011 to 2016 and –38.83% ($P<.001$) for 2016 to 2020. The age distribution of those searching for topics related to PE showed that the population aged 20 to 40 years comprised nearly 70% of the total search inquiries (second was 17.95% in the age group younger than 19 years). People from East China made over 50% of the total search queries.

Conclusions: The fluctuating online popularity of PE searches reflects the real-time population demands. It may help medical professionals better understand population interest, population concerns, regional variations, and gender differences on a nationwide scale and make disease-specific health care policies. The internet search data could be more reliable when the insufficient and lagging registry data are completed.

(J Med Internet Res 2021;23(8):e30271) doi:10.2196/30271

KEYWORDS

premature ejaculation; Baidu Index; infodemiology; public interest; patients' concern; sexuality; sexual dysfunction

Introduction

Premature ejaculation (PE) is one of the most described psychosocial stress and sexual complaint worldwide [1]. The estimated PE prevalence rate is 16% and over 20% in an internet-based survey [2,3]. Characterized with poor controlled and rapid ejaculation, the PE condition negatively impacts both sexual partners with symptomatic psychological problems and humanistic economic burdens [2,4]. However, instead of seeking help from professionals, previous reports revealed that most patients with PE are prone to initiate information-seeking behavior by using the internet due to embarrassment and wrong beliefs [5,6]. The PE prevalence and its actual impact may hence remain unveiled and underestimated.

The quality and liability of online health care–related content have been previously examined and scrutinized. Though the content was mainly acceptable, the amount of misleading content and their impact on the audience cannot be ignored. Currently, *searching before seeing a doctor* has become a trend in patients. The inquiry frequency and concerning problems from the users have been well documented by internet platforms and their demographic data. Using data from these internet platforms, infodemiology research has been successfully practiced in reporting disease incidence [7,8], surveilling pandemic outbreaks [9,10], and analyzing other public health events and related public awareness [11,12]. Previously, public interest and the change over time of the search volume in sexual dysfunction were analyzed [13,14]. It was shown that consulting *Dr Internet* influences the decision-making process of inquisitive people. Additionally, the extent of misinformation penetration could not be neglected [6,13]. However, such data in mainland China is still lacking.

By 2020, the netizens population has reached 940 million in mainland China [15]. With the 766 million users actively seeking medical information and inquiring about symptoms for diagnosis confirmation on the internet, we believe it is crucial to evaluate users search behavior and their interest or beliefs [16]. As the leading search engine in mainland China, Baidu has taken 92.1% of the search volume and 93.1% coverage of the use rate exclusively [17]. Baidu's big data analyzing platform, Baidu Index, enables users to track the popularity change and correlated demand of one specific phrase [7,8]. Therefore, we aim to evaluate the online search trend and user demand related to PE problems and the genuine needs from a *real-world* geospatial and temporal database.

Methods

Keyword Selection and Data Retrieval

This study was mainly referring to the temporal search trends of PE-related terms in Chinese. The PE keyword was identified by referring to the International Society for Sexual Medicine's definition [1]. To reduce the language habit–derived differences and biases, all possible synonyms or complex derivatives were screened and selected as previously described [8,18]. According to the available search keywords, four categories of PE topics were identified. All included PE search keywords were checked on the Baidu Index platform for their availability and are listed in Multimedia Appendix 1.

In the trend module, the Baidu Search Index (BSI) value is provided for each keyword at the national and subnational scale [8]. The BSI value is a numerical value that represents the sequential search volume on a daily basis. Hence, each keyword's monthly search index values from January 1, 2011, to December 31, 2020, were collected at the national and provincial level from the platform [8,18]. For the terms with multiple available synonyms, the searching value of each synonym was summed [8,18]. In the search-demand module, the top 10 keyword-related phrases were listed and sorted with their BSI value. The user age, gender, and region distribution data were collected from the demographic portrait module. Therefore, the user demand and user geodemographic data were collected from the Baidu database to analyze users' demand and public awareness concerning the PE issue.

Data Analysis

For each PE topic, the corresponding BSI data were plotted sequentially to describe the trend of public attention. The monthly search index of each PE topic was sorted annually, and the overall time trend change for each domain was determined by the annual percent change (APC) model. This model is designed to examine the overtime change of popularity over a specified fixed interval [19]. The APC was calculated by the Joinpoint Regression model, Program Version 4.7.0.0 (Statistical Research and Applications Branch, National Cancer Institute). A *P*<.05 was considered statistically significant.

Statistical Analysis

The database was constructed with Excel 2019 (Microsoft Corporation). We used Prism 8 for macOS, version 8.4.0 (455; GraphPad Software) to conduct statistical analysis and create figures.

Results

Web-Based Data Trends in PE Topics

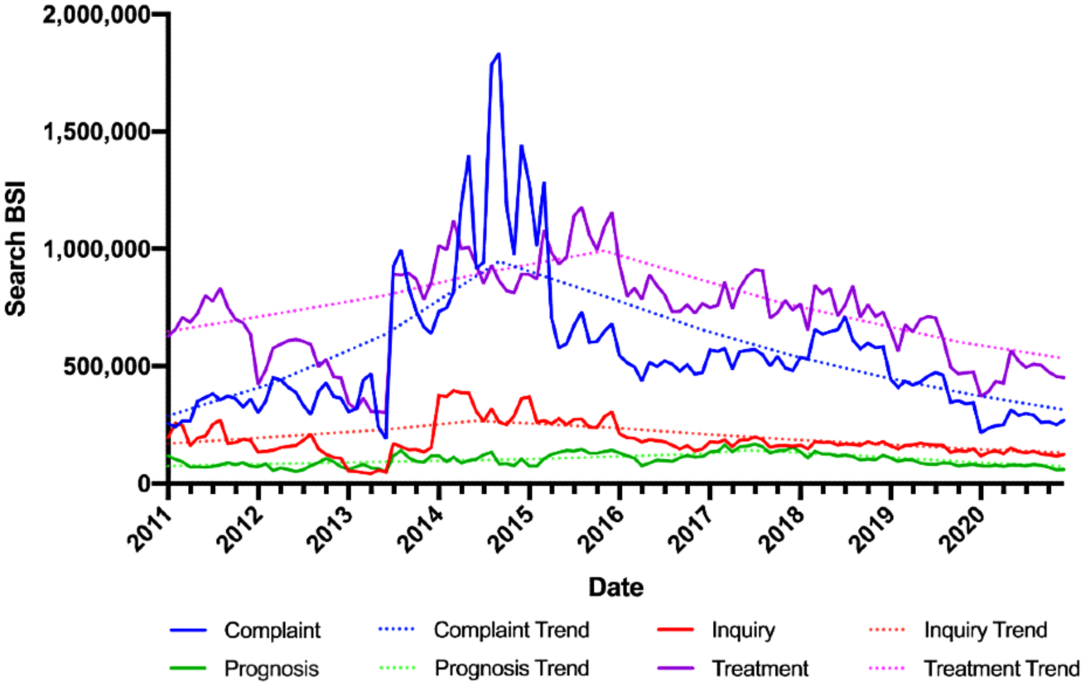
We summarized the total BSI of PE-related search keywords in the past 10 years. The retrieved 36 search keywords were manually categorized into four topics, complaint, inquiry, prognosis, and treatment, based on search keywords' content by the expert panel of medical professionals with over 5 years of experience. The total BSI value of these PE search keywords was 187,558,154. The BSI for each PE topic varied greatly. The search percentage in treatment (n=86,840,487, 46.30%) and complaint (n=66,387,459, 35.40%) accounted for the majority of the PE searches, leaving the inquiry (n=22,320,901, 11.90%) and prognosis (n=12,009,307, 6.40%) topics accounting for less than 20% of the total search (Table 1).

The monthly time series curves of the BSI for each PE topic and the APC trend lines are demonstrated in Figure 1. According to the average count of the annual BSI for each PE topic, these topic trends shared a similar increase for a period of time and redescended at the end of 2020. The following are the overall time average APC for the topics: complaint (1.18%; *P*=.94), inquiry (–2.9%; *P*=.73), prognosis (–0.32%; *P*=.91), and treatment (2.13%; *P*=.65). Specifically, the average APC for the topic complaint was 48.80% (*P*<.001) for 2011 to 2014 and –16.82% (*P*<.001) for 2014 to 2020. The average APC for the topic inquiry was 16.21% for 2011 to 2014 (*P*=.41) and –11.00% (*P*<.001) for 2014 to 2020. For the topic prognosis, the average APC was 11.18% (*P*<.001) from 2011 to 2017 and –19.86% (*P*<.001) from 2017 to 2020. For the topic treatment, the average APC was 14.04% (*P*<.001) from 2011 to 2016 and –38.83% (*P*<.001) from 2016 to 2020. The detailed search trend of the summed searches detailed in each topic are listed in Multimedia Appendix 2.

Table 1. Search popularity of each topic in premature ejaculation.

Topic	Baidu Search Index (n=187,558,154), n (%)
Complaint	66,387,459 (35.40)
Inquiry	22,320,901 (11.90)
Prognosis	12,009,307 (6.40)
Treatment	86,840,487 (46.30)

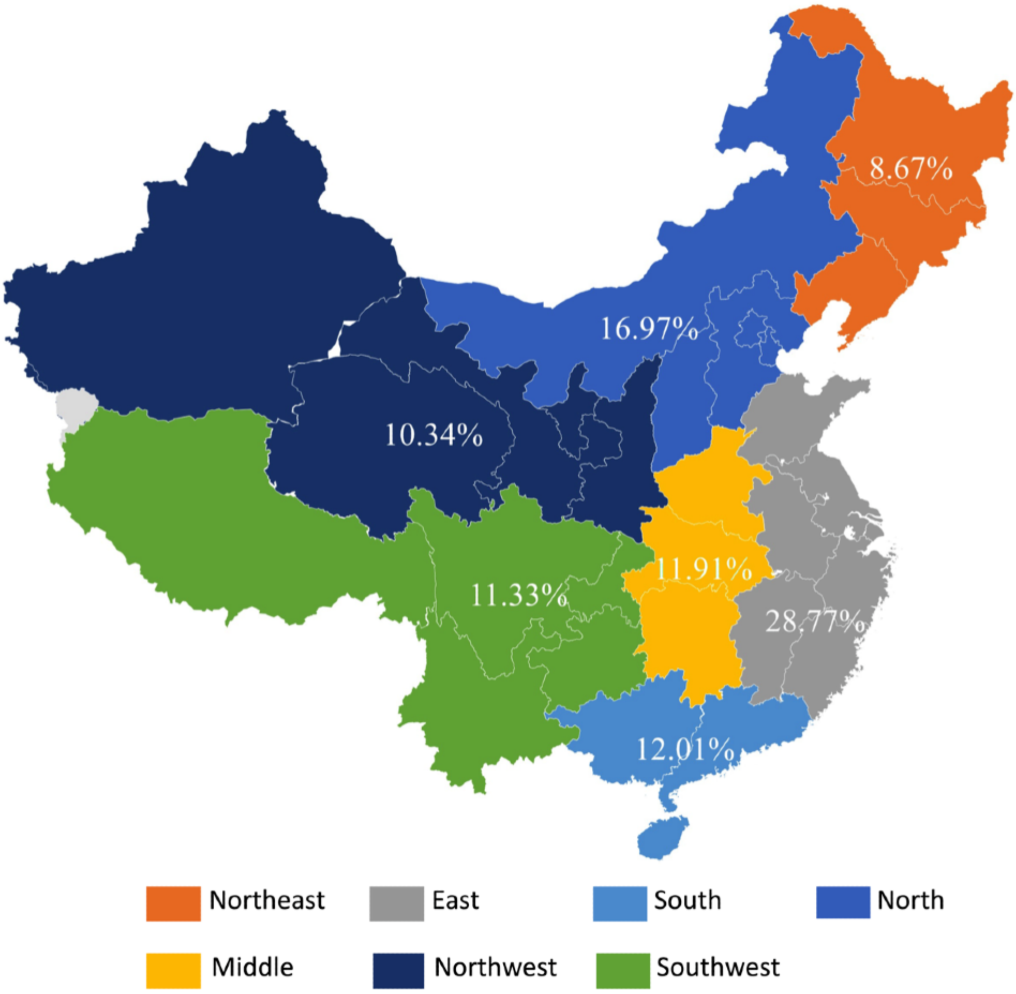
Figure 1. Online interest in premature ejaculation topics over the last 10 years. BSI: Baidu Search Index.



Geographic Differences

The PE BSI geographic distribution was calculated based on provincial data and sorted according to Chinese administrative divisions. The seven regions were as follows: Northeast China, East China, South China, North China, Central China, and Northwest and Southwest China. In Figure 2, the 10-year regional BSI proportions for all PE topics are presented on a map of mainland China with a valid searching record. Notably, people from eastern China (Northeast, East, and South China) made 57.75% (BSI: 108,317,892/187,558,154) of the total search queries. Nevertheless, the queries from west China (Northwest and Southwest China) only made 21.67% (BSI: 40,645,892/187,558,154) of search queries.

Figure 2. Regional distribution of online interest in premature ejaculation searches over the last 10 years.



Demographic Differences

From the age distribution (Table 2), in total, it was demonstrated that nearly 70.24% (BSI: 131,732,080/187,558,154) of the PE searches was inquired by the population aged 20 to 39 years (20-29 years and 30-39 years), followed by the population aged 40 to 49 years and younger than 19 years. People older than 50 years accounted for 1.87% (BSI: 3,516,547/187,558,154) of the total popularity. A similar pattern was observed in each PE topic subgroup. With regard to gender differences, the total inquiry was mainly dominated by the male group (Table 3). In detail, the treatment topic was the main concern from the male population, followed by the complaint, prognosis, and inquiry topics.

Table 2. Age percentages of search popularity in premature ejaculation topics.

Topic	Baidu Search Index, n (%)				
	≤19 years	20-29 years	30-39 years	40-49 years	≥50 years
General (n=187,558,154)	33,738,405 (17.98)	80,330,533 (42.83)	51,401,547 (27.40)	18,571,122 (9.92)	3,516,547 (1.87)
Complaint (n=66,387,459)	11,590,226 (17.46)	29,134,384 (43.88)	18,625,737 (28.06)	6,048,631 (9.11)	988,481 (1.49)
Inquiry (n=22,320,901)	4,731,811 (21.19)	10,623,644 (47.59)	5,217,263 (23.37)	1,567,679 (7.02)	180,504 (0.81)
Prognosis (n=12,009,307)	3,252,774 (27.08)	4,391,126 (36.56)	3,267,039 (27.20)	975,100 (8.12)	123,268 (1.02)
Treatment (n=86,840,487)	14,163,594 (16.31)	36,181,379 (41.66)	24,291,508 (27.97)	9,979,712 (11.49)	2,224,294 (2.56)

Table 3. Gender percentages of search popularity in premature ejaculation topics.

Topic	Female, n (%)	Male, n (%)
General (n=187,558,154)	46,556,413 (24.82)	141,001,741 (75.18)
Complaint (n=66,387,459)	20,435,182 (30.78)	45,952,277 (69.22)
Inquiry (n=22,320,901)	9,144,492 (40.97)	13,176,409 (59.03)
Prognosis (n=12,009,307)	3,928,248 (32.71)	8,081,059 (67.29)
Treatment (n=86,840,487)	13,048,492 (15.03)	73,791,995 (84.97)

Keyword-Relevant Terms and Search Frequency

In the Baidu Index platform, the user demand data is available for review by sorting the top-searched keyword-relevant terms. The terms were categorized into 14 types to clarify the users' main concerns and sort their popularity according to the content and implied intention. To better explicitly and comprehensively describe users' concerns, these categories were defined as (1) irrelevant, (2) complaint, (3) inquiry of etiology, (4) treatment, (5) health-related information, (6) diagnosis, (7) hospital and product, (8) diagnosis confirmation, (9) test and examinations, (10) prognosis, (11) traditional Chinese medicine (TCM) complaint, (12) TCM inquiry, (13) TCM regiment, and (14) TCM materials. The total BSI for the user demand terms was 1,518,298,328, which is nearly 10 times the requests for PE terms. However, only 43.74% (BSI: 664,107,756/1,518,298,328) of the demand search terms were identified as relevant to PE. Detailed distributions of these relative terms were presented in Figure 3. The categorized users' demand in each PE topic and the detailed treatment inquiries are presented in Figures 4 and 5. Furthermore, the top 3 related terms and their BSI are listed in Multimedia Appendix 3.

Figure 3. Term categories related to premature ejaculation searches in the Baidu Index demand graph. A: relevant ratio of term BSI; B: distribution of term. BSI: Baidu Search Index; TCM: traditional Chinese medicine.

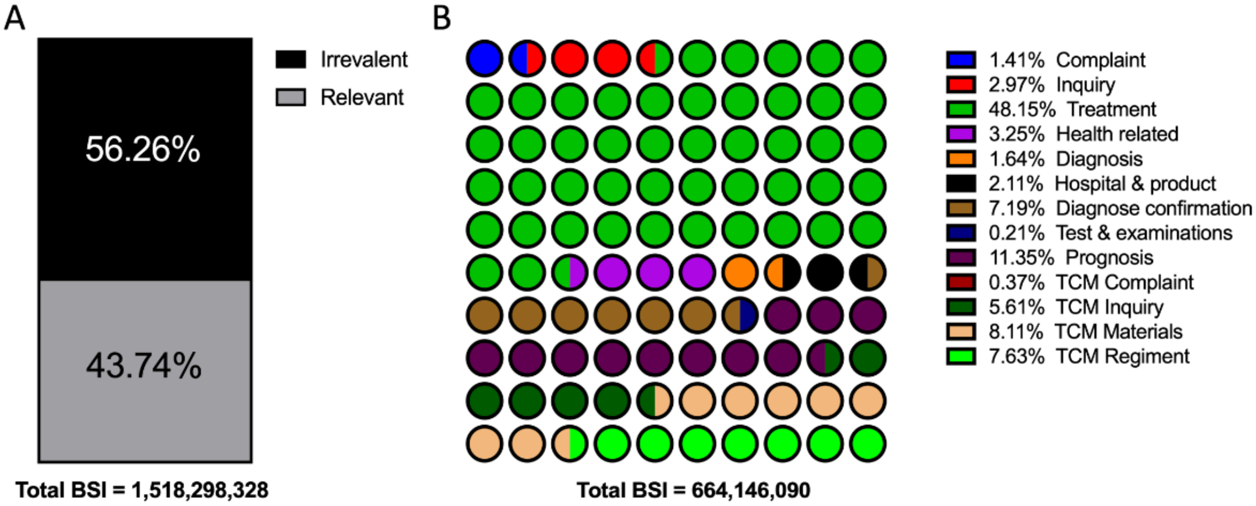


Figure 4. Term categories related to each premature ejaculation topic search in the Baidu Index demand graph. A: topic complaint; B: topic inquiry; C: topic prognosis; D: topic treatment. BSI: Baidu Search Index; TCM: traditional Chinese medicine.

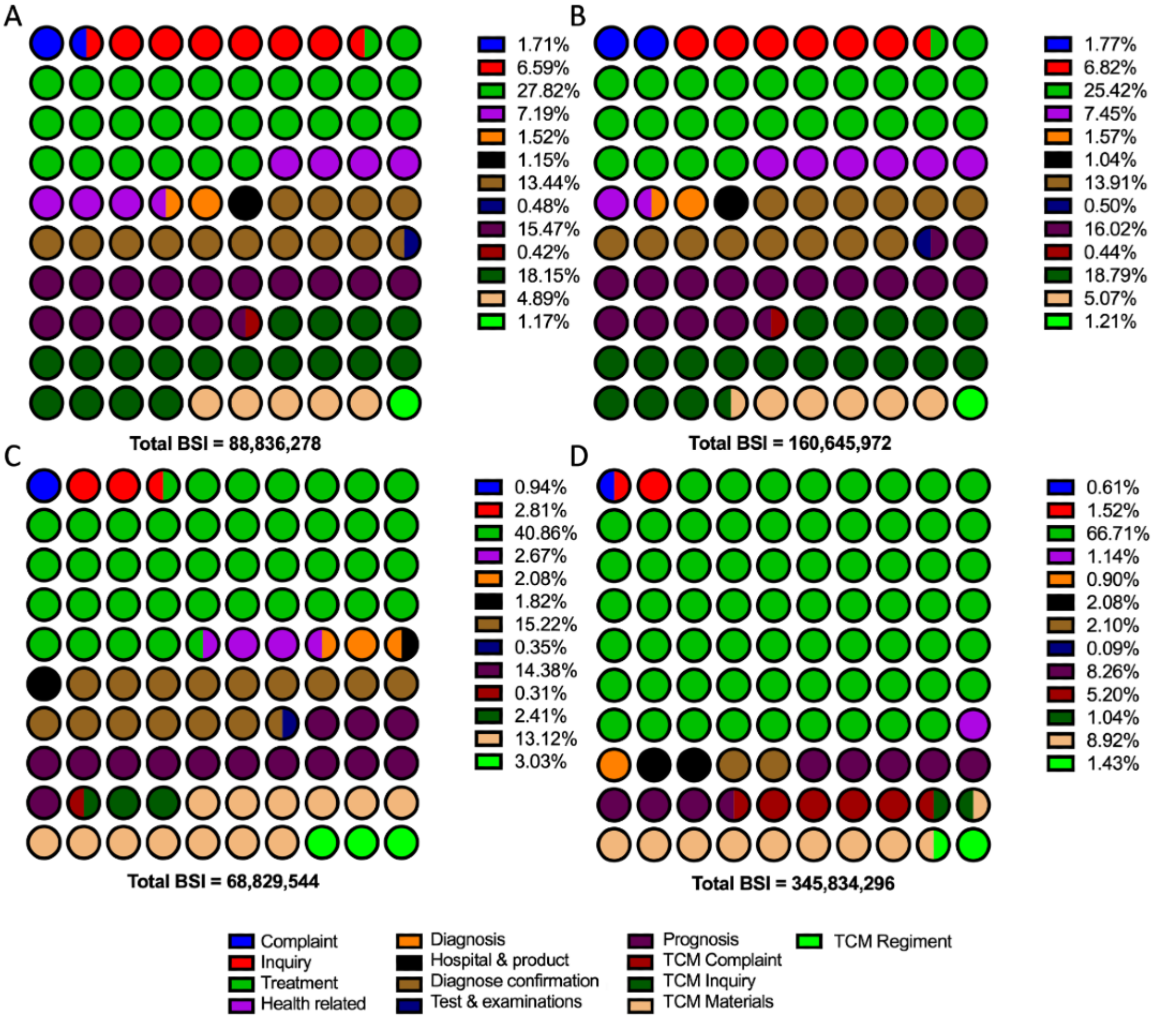
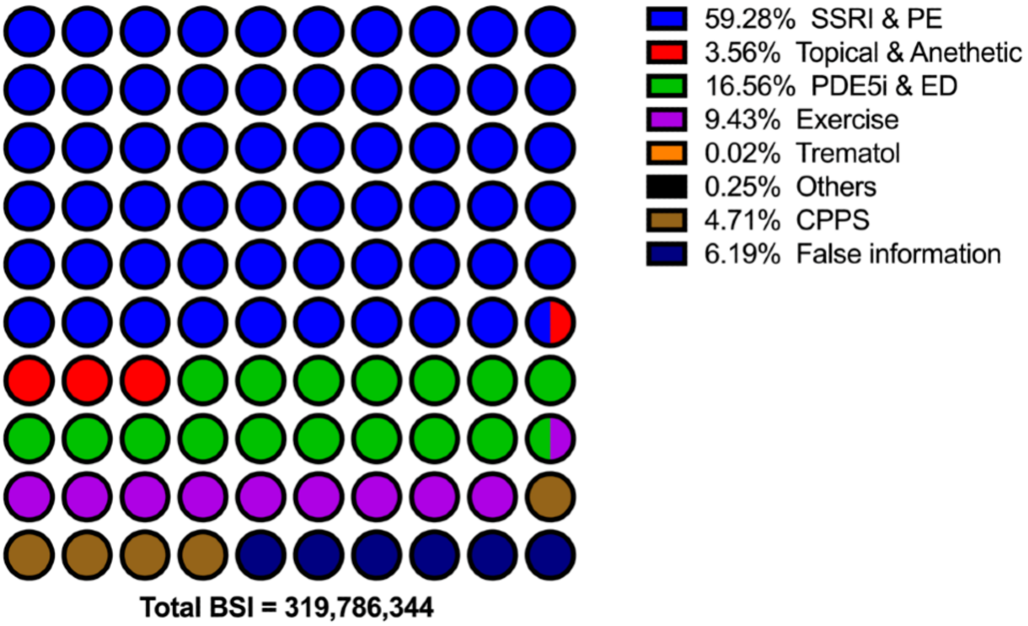


Figure 5. Classification of the detailed options in PE-related treatment search terms. BSI: Baidu Search Index; CPPS: chronic pelvic pain syndrome; ED: erectile dysfunction; PDE5i: phosphodiesterase type 5 inhibitor; PE: premature ejaculation; SSRI: selective serotonin reuptake inhibitor.



Discussion

Principal Findings

In investigating PE online interest in mainland China, 36 PE search keywords were identified as valid and could be categorized into the topics complaint, inquiry, prognosis, and treatment. Information on search trend and the geographic, demographic, and user demand information could be identified based on data from the BSI platform.

Baidu is the leading search platform in mainland China. Every day, over billions of user-initiated information queries are made on this platform. With the substantial amount of user-generated data, the medical professionals were enabled with a new insight into users' information-seeking behaviors and health care problems. Studies have confirmed that the big data from the online platform is capable of assisting with forecasting pandemic outbreaks, identifying population interests, monitoring health education campaigns, and tracking the popularity trend of patients' preferences [20-22]. Our investigation revealed that the PE search trend on the Baidu Index platform reflects the overall public concern of this health issue.

In our research, 36 PE keywords were categorized into four topic categories. Searches for the complaint topic accounted for over 35% (BSI: 66,387,459/187,558,154) of the search popularity for PE inquiries. This high proportion is probably because the PE search keywords are also words describing symptoms. When it was intuitive and simple to express without content restriction, starting a query with these terms could allow users to decide additional queries depending on the retrieved information. The treatment topic included 21 search keywords, describing PE treatment ranging from medication, surgery, or other available options. From these search keywords, the conveyed definition and concept were restricted and specific. Therefore, this result reveals that Chinese users are likely to assume they have a health problem and are more concerned about the availability and options of solving this problem. Despite the lack of basic understanding of this disease, such as its definition, morbidity, etiology, pathology, or diagnostic criteria, the treatment options were users' primary concern in health issues like PE. Though six search keywords concerned PE definition and criteria, the total search popularity of the topic inquiry accounted for only 11.90% (BSI: 22,320,901/187,558,154). Hence, a high rate of online self-diagnosis and treatment was observed. Internet self-diagnosis has caused concern about diagnosis delay and misleading treatments [23]. With unproven or biased information that is frequently presented on the internet, the population is given unrealistic expectations. A distrust of professionals with "objective" reasoning could also exacerbate their anxiety about their own problems [24]. Our findings partly represent these issues. Therefore, improving internet health care and optimizing medicine services relies on elevating service quality and popularizing basic medical concepts [25].

When evaluating the popularity trend, a similar pattern was observed for the four PE topics. For each topic, the sustained growing trends were only observed in the beginning years. This trend patten reveals that, although the PE incident rate in the population may remain constant, the users' medical advice-seeking habits may have changed. The medical information from the Baidu platform has always been in public dispute. As the most commonly used search engine in mainland China, Baidu has long been criticized for presenting the top-ranked sites based on higher bids [23]. The biased and inductive information with higher bids are always listed at the top, making high-quality or evidence-based medical information from universities or medical centers more time consuming to obtain [23]. When the Wei Zexi incident broke out in 2016, public critiques of Baidu had reached a peak because the top listed and false advertised treatments claiming to cure synovial sarcoma from this search engine were deemed the leading cause for a 21-year-old student's death [26]. In 2017, Hu et al [23] found that most health information online from the first 100 links for keywords were generally inadequate. Furthermore, from a recent investigation by Chen et al [27], the content available on Baidu were less credible and inferior to those from Google. Consequently, instead of serving as a search engine that presents information on health education and credible content, Baidu more resembles an advertising platform. Therefore, local public health authorities should participate more and make regulations on the internet service industry, encouraging them to provide more qualified and well-produced information, unbiased information, and evidence-based recommendations on their platform. Hence, the trends in our research indicate that, although the PE prevalence may remain stable, the public awareness may change as people change their information-seeking behavior in response to poor quality content, impacting search popularity.

From the geographic data, we discovered that the PE search popularity was dominated by searches from the East China, North China, and South China regions, where the top four Chinese megacities (Beijing, Shanghai, Guangzhou, and Shenzhen) are located. Although national PE epidemiology data are lacking, this geographic distribution was in line with the leading regional economic level and the demographic distribution [28]. Coinciding with the finding from Xu et al [8] and Wang et al [7], the search popularity's interregional gap may also be a result of the differences in socioeconomic status, population structure, healthy concepts, and health care policies [9]. Medical professionals could gain insight into the public concern toward PE and customize their clinical session accordingly.

PE was reported as the most common sexual complaint in males in all age groups in previous epidemiological studies [29]. Though it has long been believed that the PE incidence is correlated with age differences, the existing age PE prevalence data were controversial [30-32]. In our investigation, over 70% of PE online search queries came from males aged 20 to 39 years. This result appears to contradict the Global Online Sexuality Survey, from which the peak PE prevalence was observed in men in their forties [32]. Nevertheless, from an internet-based survey, the highest self-reporting PE incident was for those in their thirties and twenties [30]. The discrepancy between these previous research results may be caused by the screening method, diagnostic standard, and inclusion criteria of risk factors such as age, coitus regularity, circumcision, and masturbation history [31,32]. We believe the openness in Son et al’s [30] investigation resembles the online inquiry on a searching platform. Their discoveries were similar to ours, the results presenting the natural concerns of users and the related popularity distribution [30]. In addition, we noticed the PE inquiry from the female population had reached 24.82% (BSI: 46,556,413) in total and 40.97% (BSI: 9,144,492) in the topic inquiry. Previously, the investigated partners’ satisfaction rates were 45% from the individual with PE’s perspective [33]. According to Zhang et al [34], although only 21.39% of females tended to report PE as hampering sex satisfaction, females are more adversely experienced with relationship issues, such as arousal difficulty, weak sexual desire, and organism inability [34]. Though the cultural differences may underlie searching preference, as the tip of an iceberg, our results may show the concerns and demands of PE issues troubling females. Hence, PE impact on females should be stressed.

In the user demand section, the top 10 most related terms were used for each keyword search. From this section, practitioners were able to gain insight into the patients’ most concerning problems and to confirm individuals’ main intentions from these weekly updated data. Generally, these related terms reflected the wide-ranging scope of concerns when searching PE-related content, such as PE comorbidity, PE training practices, PE etiology, and reproductive influences.

We noticed that the most popular term was “The fastest and easiest way to cure PE?” The high prevalence of this term echoes the topic popularity finding in our research, indicating that the users’ most concerning issue is related to problem solving. Hence, users’ self-diagnosis and self-treatment of issues should be stressed.

The relationship between PE and chronic prostatitis (CP) has been long observed [35]. Suggested as an organic PE cause, the existence of CP in PE is yet to be fully understood. The sexual dysfunction prevalence in males with CP has reached 49% in previous investigations, and the severity of the symptom scoring between these two clinic complaints is well correlated [36,37]. Supporting these investigations, we discovered that “prostate” and “prostatitis” were the most inquired terms from users. This popularity revealed that prostatic problems are the primary concern for patients with PE; also, these people may tend to self-diagnose themselves with prostatitis instead of ejaculatory dysfunction. Whether this concern results from the doctor’s instruction or is based on the patient with PE’s complaint, it is necessary to screen for PE and CP complications.

Additionally, in investigating the users’ demands in treatment, the most queried terms were related to selective serotonin reuptake inhibitors. The followed treatment regimens were phosphodiesterase type 5 inhibitor, topical anesthetic agents, behavior therapy, and tramadol, accounting for over 80% of the search popularity, and these treatment options were either recommended by the sexual dysfunction treatment guidelines or evaluated for safety and efficacy [38,39]. This information could alert the patients to treatment avoidance and contraindication. In addition, better compliance could facilitate patients’ comprehension of the doctor’s instructions.

Nevertheless, terms with false information and their popularity are not noteworthy. These terms typically suggest applying homemade regiments from everyday domestic items on to the penis, ranging from vegetable mixtures with ginger, leeks, or onions to toothpaste or the *Fengyou essence* (a type of essential balm). Though patients could not be harmed much from the vegetable mix remedies aside from allergies or mucosal lesions, damages from the toothpaste and the *Fengyou essence* can be more consequential [6], for the main ingredients of *Fengyou essence* and toothpaste are more abrasive and irritative pharmaceutical materials; the improper application of these regimens with the wrong dose could cause severe skin abrasion or even damage the genital organ [6,40].

Some limitations must be addressed in this study. First, the Baidu Index is only available for search data on the Baidu platform. Though Baidu is the most widely used platform and monopolizes the searching requests in mainland China, the users’ shift to social media for searching is rising. Therefore, lacking a uniform recording system, data from these growing social media sites are not able to be incorporated or be assessed without bias . In addition, in considering the user privacy security, only the user age, gender, and living region information was available for the demographic research. For an in-depth demographic investigation, data such as socioeconomic status, ethnicity, or educational background are required. Additionally, instead of representing the actual search frequency, the BSI is just a weighted index derivative and cannot manifest real-world disease prevalence. Nevertheless, before a nationally conducted PE prevalence and demographic investigation is available, infodemiology research is practical to investigate the PE problems and users’ information-seeking behaviors from the population perspective.

To our best knowledge, this is the first infodemiology study that investigates PE public concerns in mainland China. We chose to examine the PE popularity because its clinical manifestations are intuitive and easy to describe. More importantly, patients with sexual dysfunction are willing to search online due to privacy concerns. Hence, the real-time updated searching data helps to improve practice standards and policy making for medical professionals and health officials.

Conclusion

The fluctuating online popularity of PE searches reflects the real-time population demands. It may help medical professionals better understand population interest, population concerns, regional variations, and gender differences on a nationwide scale and make disease-specific health care policies. The internet search data could be more reliable when the insufficient and lagging registry data are completed.

Acknowledgments

This paper was supported by National Natural Science Foundation of China (No. 81800666 and 81670688), Natural Science Foundation of Guangdong Province (No.2016A030313192, 2021A1515010354, and 2018A0303130330), Technology Planning Project of Guangzhou City (No. 202103000035), Fundamental Research Funds for the Central Universities (No.19ykpy33), and the Science and Technology Achievement Transformation Foundation and 1.3.5 project for disciplines of excellence, Sichuan University West China Hospital.

Authors' Contributions

This study was conceptualized by SW and XZ. The methodology was designed by SW, MM, and XW. The investigation was carried out by XW, CW, GZ, and SW. SW and MM wrote the original draft. SW, MM, and XZ reviewed and edited the draft. Funding was acquired by XZ. CW and XZ contributed resources, and XZ supervised the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
List of searching keywords used in compositing search index.
[PDF File (Adobe PDF File), 107 KB - jmir_v23i8e30271_app1.pdf]

Multimedia Appendix 2
Supplementary figures of summed search trend and detailed search trend of the premature ejaculation search keywords.
[PDF File (Adobe PDF File), 185 KB - jmir_v23i8e30271_app2.pdf]

Multimedia Appendix 3
Top 3 terms of users' demands in premature ejaculation.
[PDF File (Adobe PDF File), 179 KB - jmir_v23i8e30271_app3.pdf]

References

1. Serefoglu EC, McMahon CG, Waldinger MD, Althof SE, Shindel A, Adaikan G, et al. An evidence-based unified definition of lifelong and acquired premature ejaculation: report of the second international society for sexual medicine ad hoc committee for the definition of premature ejaculation. Sex Med 2014 Jun;2(2):41-59 [FREE Full text] [doi: 10.1002/sm2.27] [Medline: 25356301]
2. Porst H, Montorsi F, Rosen RC, Gaynor L, Grupe S, Alexander J. The Premature Ejaculation Prevalence and Attitudes (PEPA) survey: prevalence, comorbidities, and professional help-seeking. Eur Urol 2007 Mar;51(3):816-23; discussion 824. [doi: 10.1016/j.eururo.2006.07.004] [Medline: 16934919]
3. Pozzi E, Capogrosso P, Boeri L, Cazzaniga W, Matloob R, Ventimiglia E, et al. Trends in reported male sexual dysfunction over the past decade: an evolving landscape. Int J Impot Res 2020 Jul 01:1. [doi: 10.1038/s41443-020-0324-7] [Medline: 32612274]
4. Balon R. Burden of sexual dysfunction. J Sex Marital Ther 2015 Dec 18;43(1):49-55. [doi: 10.1080/0092623x.2015.1113597]
5. LaValley SA, Kiviniemi MT, Gage-Bouchard EA. Where people look for online health information. Health Info Libr J 2017 Jun;34(2):146-155. [doi: 10.1111/hir.12143] [Medline: 27207817]
6. Gul M, Diri MA. YouTube as a source of information about premature ejaculation treatment. J Sex Med 2019 Nov;16(11):1734-1740. [doi: 10.1016/j.jsxm.2019.08.008] [Medline: 31522984]
7. Wang T, Xia Q, Chen X, Jin X. Use of Baidu Index to track Chinese online behavior and interest in kidney stones. Risk Manag Healthc Policy 2020;13:705-712. [doi: 10.2147/RMHP.S245822] [Medline: 32694922]
8. Xu C, Wang Y, Yang H, Hou J, Sun L, Zhang X, et al. Association between cancer incidence and mortality in web-based data in China: infodemiology study. J Med Internet Res 2019 Jan 29;21(1):e10677 [FREE Full text] [doi: 10.2196/10677] [Medline: 30694203]
9. Hou Z, Du F, Zhou X, Jiang H, Martin S, Larson H, et al. Cross-country comparison of public awareness, rumors, and behavioral responses to the COVID-19 epidemic: infodemiology study. J Med Internet Res 2020 Aug 03;22(8):e21143 [FREE Full text] [doi: 10.2196/21143] [Medline: 32701460]
10. Wang Q, Dong W, Yang K, Ren Z, Huang D, Zhang P, et al. Temporal and spatial analysis of COVID-19 transmission in China and its influencing factors. Int J Infect Dis 2021 Apr;105:675-685 [FREE Full text] [doi: 10.1016/j.ijid.2021.03.014] [Medline: 33711521]
11. Sousa-Pinto B, Anto A, Czarlewski W, Anto JM, Fonseca JA, Bousquet J. Assessment of the impact of media coverage on COVID-19-related Google trends data: infodemiology study. J Med Internet Res 2020 Aug 10;22(8):e19611 [FREE Full text] [doi: 10.2196/19611] [Medline: 32530816]

12. Cacciamani GE, Sebben M, Tafuri A, Nassiri N, Cocci A, Russo GI, et al. Consulting 'Dr. Google' for minimally invasive urological oncological surgeries: a contemporary web-based trend analysis. Int J Med Robot 2021 Aug;17(4):e2250. [doi: 10.1002/rcs.2250] [Medline: 33667326]
13. Davis NF, Smyth LG, Flood HD. Detecting internet activity for erectile dysfunction using search engine query data in the Republic of Ireland. BJU Int 2012 Dec;110(11 Pt C):E939-E942. [doi: 10.1111/j.1464-410X.2012.11237.x] [Medline: 22587462]
14. Russo GI, di Mauro M, Cocci A, Cacciamani G, Cimino S, Serefoglu EC, EAU-YAU Men’s Health Working Group. Consulting "Dr Google" for sexual dysfunction: a contemporary worldwide trend analysis. Int J Impot Res 2020 Jul;32(4):455-461. [doi: 10.1038/s41443-019-0203-2] [Medline: 31591474]
15. The 46th China Statistical Report on Internet Development. China Internet Network Information Center. 2020. URL: https://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg/ [accessed 2021-07-29]
16. Chinese netizens' demand on science search, popularization behavior report (first quarter 2019). China Association for Science and Technology. 2019. URL: https://www.cast.org.cn/art/2019/4/26/art_1281_94546.html [accessed 2021-07-29]
17. Clement J. Worldwide desktop market share of leading search engines from January 2010 to July 2020. Statista. 2020. URL: https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/ [accessed 2021-07-29]
18. Masui K, Yoshimura K, Soda T, Kamba T, Kanematsu A, Nakamura E. [Health-care seeking behavior related to lower urinary tract symptoms : a population-based study in Japan]. Hinyokika Kiyo 2009 Nov;55(11):685-689 [FREE Full text] [Medline: 19946185]
19. Ho L, Chu PS, Consigliere DT, Zainuddin ZM, Bolong D, Chan C, et al. Symptom prevalence, bother, and treatment satisfaction in men with lower urinary tract symptoms in Southeast Asia: a multinational, cross-sectional survey. World J Urol 2018 Jan;36(1):79-86. [doi: 10.1007/s00345-017-2097-2] [Medline: 29051978]
20. Li F, Li M, Guan P, Ma S, Cui L. Mapping publication trends and identifying hot spots of research on Internet health information seeking behavior: a quantitative and co-word biclustering analysis. J Med Internet Res 2015 Mar 25;17(3):e81 [FREE Full text] [doi: 10.2196/jmir.3326] [Medline: 25830358]
21. Mayo-Yáñez M, Calvo-Henríquez C, Chiesa-Estomba C, Lechien JR, González-Torres L. Google Trends application for the study of information search behaviour on oropharyngeal cancer in Spain. Eur Arch Otorhinolaryngol 2021 Jul;278(7):2569-2575. [doi: 10.1007/s00405-020-06494-7] [Medline: 33237476]
22. 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]
23. Hu X, Pan H, He W, Hua H, Yan Z. Evaluation of the content quality of websites for recurrent aphthous ulcers and oral lichen planus. BMC Oral Health 2017 Dec 29;17(1):170 [FREE Full text] [doi: 10.1186/s12903-017-0467-1] [Medline: 29284468]
24. Lewis LF. Exploring the experience of self-diagnosis of autism spectrum disorder in adults. Arch Psychiatr Nurs 2016 Oct;30(5):575-580. [doi: 10.1016/j.apnu.2016.03.009] [Medline: 27654240]
25. Robertson N, Polonsky M, McQuilken L. Are my symptoms serious Dr Google? A resource-based typology of value co-destruction in online self-diagnosis. Australasian Marketing J 2021 Jan 27;22(3):246-256. [doi: 10.1016/j.ausmj.2014.08.009]
26. Pogacar FA, Ghenai A, Smucker MD, Clarke CLA. The positive and negative influence of search results on people's decisions about the efficacy of medical treatments. In: Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval. 2017 Oct 1 Presented at: ICTIR '17; October 1-4, 2017; Amsterdam, The Netherlands. [doi: 10.1145/3121050.3121074]
27. Chen T, Gentry S, Qiu D, Deng Y, Notley C, Cheng G, et al. Online information on electronic cigarettes: comparative study of relevant websites from Baidu and Google Search engines. J Med Internet Res 2020 Jan 24;22(1):e14725 [FREE Full text] [doi: 10.2196/14725] [Medline: 32012069]
28. Wang W, Guo Y, Zhang D, Tian Y, Zhang X. The prevalence of benign prostatic hyperplasia in mainland China: evidence from epidemiological surveys. Sci Rep 2015 Aug 26;5:13546. [doi: 10.1038/srep13546] [Medline: 26306721]
29. Althof SE. Prevalence, characteristics and implications of premature ejaculation/rapid ejaculation. J Urol 2006 Mar;175(3 Pt 1):842-848. [doi: 10.1016/S0022-5347(05)00341-1] [Medline: 16469562]
30. Son H, Song SH, Kim SW, Paick J. Self-reported premature ejaculation prevalence and characteristics in Korean young males: community-based data from an internet survey. J Androl 2010;31(6):540-546. [doi: 10.2164/jandrol.110.010355] [Medline: 20671139]
31. Song SH, Choi WS, Son H, Paick J. Validity of the Premature Ejaculation Diagnostic Tool in four subgroups of premature ejaculation syndrome: data from the Korean Internet Sexuality Survey - part 1. Sex Health 2014 Mar;11(1):73-80. [doi: 10.1071/SH13171] [Medline: 24670301]
32. Shaeer O. The global online sexuality survey (GOSS): the United States of America in 2011 Chapter III--Premature ejaculation among English-speaking male Internet users. J Sex Med 2013 Jul;10(7):1882-1888. [doi: 10.1111/jsm.12187] [Medline: 23668379]

33. Waldinger MD, Quinn P, Dilleen M, Mundayat R, Schweitzer DH, Boolell M. A multinational population survey of intravaginal ejaculation latency time. J Sex Med 2005 Jul;2(4):492-497. [doi: 10.1111/j.1743-6109.2005.00070.x] [Medline: 16422843]
34. Zhang J, Li F, Li H, Zhang Z, Yang B, Li H. Clinical features of and couple's attitudes towards premature ejaculation: a multicenter cross-sectional study. Aging Male 2020 Dec;23(5):946-952. [doi: 10.1080/13685538.2019.1640194] [Medline: 31306035]
35. McMahon CG, Jannini EA, Serefoglu EC, Hellstrom WJG. The pathophysiology of acquired premature ejaculation. Transl Androl Urol 2016 Aug;5(4):434-449. [doi: 10.21037/tau.2016.07.06] [Medline: 27652216]
36. Liang C, Zhang X, Hao Z, Shi H, Wang K. Prevalence of sexual dysfunction in Chinese men with chronic prostatitis. BJU Int 2004 Mar;93(4):568-570. [doi: 10.1111/j.1464-410x.2003.04662.x] [Medline: 15008731]
37. Lee JH, Lee SW. Relationship between premature ejaculation and chronic prostatitis/chronic pelvic pain syndrome. J Sex Med 2015 Mar;12(3):697-704. [doi: 10.1111/jsm.12796] [Medline: 25475760]
38. Eassa BI, El-Shazly MA. Safety and efficacy of tramadol hydrochloride on treatment of premature ejaculation. Asian J Androl 2013 Jan;15(1):138-142 [FREE Full text] [doi: 10.1038/aja.2012.96] [Medline: 23103596]
39. Jiang M, Yan G, Deng H, Liang H, Lin Y, Zhang X. The efficacy of regular penis-root masturbation, versus Kegel exercise in the treatment of primary premature ejaculation: A quasi-randomised controlled trial. Andrologia 2020 Mar;52(1):e13473. [doi: 10.1111/and.13473] [Medline: 31746051]
40. Wen Y, Ma T, Chen X, Liu Z, Zhu C, Zhang Y, et al. Essential balm: a strong repellent against foraging and defending red imported fire ants (hymenoptera: formicidae). J Econ Entomol 2016 Aug;109(4):1827-1833. [doi: 10.1093/jee/tow130] [Medline: 27298425]

Abbreviations

APC: annual percent change
BSI: Baidu Search Index
CP: chronic prostatitis
PE: premature ejaculation
TCM: traditional Chinese medicine

Edited by R Kukafka; submitted 07.05.21; peer-reviewed by H Ma, B Yang, M Wang; comments to author 28.06.21; revised version received 05.07.21; accepted 05.07.21; published 26.08.21.

Please cite as:
Wei S, Ma M, Wen X, Wu C, Zhu G, Zhou X
Online Public Attention Toward Premature Ejaculation in Mainland China: Infodemiology Study Using the Baidu Index
J Med Internet Res 2021;23(8):e30271
URL: https://www.jmir.org/2021/8/e30271
doi:10.2196/30271
PMID:34435970

©Shanzun Wei, Ming Ma, Xi Wen, Changjing Wu, Guonian Zhu, Xiangfu Zhou. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Exploring the Expression Differences Between Professionals and Laypeople Toward the COVID-19 Vaccine: Text Mining Approach

Chen Luo[1,2*], MA; Kaiyuan Ji[1*], MA; Yulong Tang[3], MA; Zhiyuan Du[1], BE

[1]School of Journalism and Communication, Tsinghua University, Beijing, China
[2]The Faculty of International Media, Communication University of China, Beijing, China
[3]Institute of Communication Studies, Communication University of China, Beijing, China
[*]these authors contributed equally

Corresponding Author:
Yulong Tang, MA
Institute of Communication Studies
Communication University of China
No 1 Dingfuzhuang East Street, Chaoyang District
Beijing, 100024
China
Phone: 86 13217810927
Email: longbao0927@163.com

Abstract

Background: COVID-19 is still rampant all over the world. Until now, the COVID-19 vaccine is the most promising measure to subdue contagion and achieve herd immunity. However, public vaccination intention is suboptimal. A clear division lies between medical professionals and laypeople. While most professionals eagerly promote the vaccination campaign, some laypeople exude suspicion, hesitancy, and even opposition toward COVID-19 vaccines.

Objective: This study aims to employ a text mining approach to examine expression differences and thematic disparities between the professionals and laypeople within the COVID-19 vaccine context.

Methods: We collected 3196 answers under 65 filtered questions concerning the COVID-19 vaccine from the China-based question and answer forum Zhihu. The questions were classified into 5 categories depending on their contents and description: adverse reactions, vaccination, vaccine effectiveness, social implications of vaccine, and vaccine development. Respondents were also manually coded into two groups: professional and laypeople. Automated text analysis was performed to calculate fundamental expression characteristics of the 2 groups, including answer length, attitude distribution, and high-frequency words. Furthermore, structural topic modeling (STM), as a cutting-edge branch in the topic modeling family, was used to extract topics under each question category, and thematic disparities were evaluated between the 2 groups.

Results: Laypeople are more prevailing in the COVID-19 vaccine–related discussion. Regarding differences in expression characteristics, the professionals posted longer answers and showed a conservative stance toward vaccine effectiveness than did laypeople. Laypeople mentioned countries more frequently, while professionals were inclined to raise medical jargon. STM discloses prominent topics under each question category. Statistical analysis revealed that laypeople preferred the “safety of Chinese-made vaccine” topic and other vaccine-related issues in other countries. However, the professionals paid more attention to medical principles and professional standards underlying the COVID-19 vaccine. With respect to topics associated with the social implications of vaccines, the 2 groups showed no significant difference.

Conclusions: Our findings indicate that laypeople and professionals share some common grounds but also hold divergent focuses toward the COVID-19 vaccine issue. These incongruities can be summarized as “qualitatively different” in perspective rather than “quantitatively different” in scientific knowledge. Among those questions closely associated with medical expertise, the “qualitatively different” characteristic is quite conspicuous. This study boosts the current understanding of how the public perceives the COVID-19 vaccine, in a more nuanced way. Web-based question and answer forums are a bonanza for examining perception discrepancies among various identities. STM further exhibits unique strengths over the traditional topic modeling method in statistically testing the topic preference of diverse groups. Public health practitioners should be keenly aware of the cognitive differences between professionals and laypeople, and pay special attention to the topics with significant inconsistency across groups to build consensus and promote vaccination effectively.

(J Med Internet Res 2021;23(8):e30715) doi:10.2196/30715

KEYWORDS

COVID-19; vaccine; Zhihu; structural topic modeling; medical professional; laypeople; adverse reactions; vaccination; vaccine effectiveness; vaccine development

Introduction

Background

As of April 23, 2021, over 0.14 billion confirmed cases of COVID-19 and nearly 3.1 million deaths have been reported worldwide [1]. The COVID-19 vaccine has been acknowledged as one of the most effective strategies to contain the ongoing public health predicament [2]. However, what needs to be recognized is that the COVID-19 vaccine still requires cautious validation of efficacy and adverse reactions since it is a relatively innovative therapeutic intervention in development [3,4]. Owing to the intrinsic uncertainty, vaccine hesitancy and vaccine-related misinformation pervaded during the COVID-19 vaccination process [5]. Some nationwide and transnational surveys also revealed that the public’s COVID-19 vaccination intentions were suboptimal [6-8]. While numerous medical professionals have devoted themselves to vaccine development at a breakneck speed [9] and eagerly promote the massive vaccination campaign, a considerable number of laypeople expressed concerns, hesitancy, and even antagonism toward COVID-19 vaccines [5]. For instance, a recent web-based poll conducted on Twitter disclosed that more than half of the respondents doubted the safety of COVID-19 vaccines [10]. To obtain a deeper insight into the different perceptions between the professionals and laypeople toward the COVID-19 vaccine, the present study endeavors to seek the potential differentiated expressions by adopting a text mining approach on a Chinese social media platform.

The Internet as a Pivotal Communication Space for Health-Related Issues

Web-based communication provides easy and cost-effective access to a broad audience and enables interactivity and collaborative content-sharing [11]. During the past decades, the world witnessed a drastic increase in health information on the internet, along with a pronounced tendency that both patients and caregivers are growing more likely to seek health information on the internet [12]. In the meantime, people are prone to discuss health-related issues in this virtual sphere, especially during a public health crisis [13]. For example, during the COVID-19 era, some people disclosed their disease status on the internet for help-seeking [14], and a more substantial number of people talked about their own and others' symptoms as a mere natural reaction to the threat of illness [15]. Given those features, various internet platforms serve as fertile grounds for examining the public’s perceptions of health issues or events [16]. This holds true for the vaccine issue because vaccines and vaccination are buzz topics on the internet and are encompassed by provaccine and antivaccine discourses [17,18].

Recognizing the salient characteristics of the internet, health professionals spare extensive attention to utilizing the internet to launch health campaigns, deliver health knowledge, and promote behavioral change [11]. Previous studies have summarized relevant experiences in delivering health care and health interventions with the strength of internet technologies. One representative example is that some scholars classified social media into 10 categories and put forward 4 guidelines for medical professionals to better engage in web-based health communication [19]. In reality, a series of public health institutions have implemented web-based communication strategies. The Centers for Disease Control and Prevention in the United States adopted Twitter to disseminate information, interact with the audience, and alert the public throughout the Zika epidemic [20]. In a similar vein, public health agencies in Singapore also made use of Facebook for outbreak communication and communicating the Zika epidemic strategically [21]. To cope with the COVID-19 threat, many public health agencies use social media accounts to rapidly disseminate risk messages to the public to curb contagion [22]. Except for health institutions, many medical professionals practice web-based health communication spontaneously; for instance, some doctors have joined eHealth communities to exchange medical information with patients or peers [23].

Taken together, searching and exchanging health information on the internet are common phenomena nowadays; both professionals and laypeople are critical actors in the web-based health communication environment. Since the internet has prominent advantages, including low cost, easy access, broad reach, and interactivity, it facilitates the lay public to share their health concerns, seek support, enhance their health-related knowledge, and communicate with one another. Meanwhile, professionals can develop health education and interventions on the internet. For public health researchers, diversified internet platforms can be exploited to investigate varying perceptions and expressions toward various kinds of health-related issues, especially emergent ones.

Professionals vs Laypeople in Perceiving Health-Related Issues

An entrenched thought toward the divergence between professionals and laypeople emphasizes the knowledge chasm, which retains an inherent assumption that the laypeople lag behind professionals in their knowledge levels. A professional is always defined as someone who procured special knowledge or skills of a particular subject through deliberate training and practice, while laypeople usually lack formal training or practical experience [24]. Furthermore, an extended viewpoint believes that professionals’ judgments and perceptions are more objective and reliable than those of laypeople [24]. In health communication, we particularly underscore 2 additional significant dimensions stemming from the knowledge level disparity when discussing differences between professionals and laypeople: risk perception and attitude.

As a vaccine shrouded in uncertainty, societies worldwide are deluged with suspicions and debates about the COVID-19 vaccine's safety [25]. All concerns are closely connected to risk perception, which denotes people's subjective assessment of a risk's characteristics and severity [26]. Risk perception is a compound of scientific judgment and subjective factors [27]. When it comes to differences in risk perception between professionals and laypeople, one school of thought holds that owing to the differences in knowledge reservation and established mindsets, professionals usually treat risks and uncertainties from an analytical, objective, and rational perspective. Laypeople, however, are favored to rely on hypothetical, subjective, and emotional cues when perceiving risks [28-30]. Moreover, laypeople are accustomed to amplifying risks and more susceptible to psychological factors, while professionals may underestimate the dangers and accentuate the benefits of certain controversial technologies [30,31]. Another school of thought refutes those assertions by demonstrating that professionals and laypeople are unanimously influenced by emotions, worldviews, and values when forming opinions about controversial issues [27,32]. For some medical topics, the scientific literacy advantage of professionals is not more prominent than that of laypeople [33,34].

Another dimension is attitude. According to the knowledge deficit model, the lay public's skepticism toward innovative technologies can be attributed to their deficiency in scientific knowledge [35,36]. Besides, this model hypothesizes that the laypeople's and professionals' divergent opinions on the same issue can be ascribed to the public's insufficient issue-specific knowledge [37]. Therefore, a more supportive attitude toward emerging technologies could be realized by enhancing the public's scientific knowledge level or the so-called scientific literacy [35,38,39]. Although this model has been criticized by a series of empirical studies [40], it still influences health communication and science communication research. Recently, a study on emergency medicine influencers' Twitter use during the COVID-19 pandemic disclosed that medicine influencers' messages contain words with more positive and neutral emotion than those of the general public. The influencer group also has a manifest topic preference for clinical information and COVID-19 news [41].

Using Social Media to Explore Expression Differences

As one of the most burgeoning branches of internet technologies, social media has been invested with plentiful unobtrusive and naturalistic data [42,43], which makes it suitable for examining heterogeneous discussions and perceptions toward specific health-related topics or events [16]. For instance, some pundits employed tweets to gain insights and knowledge of how people discuss the human papillomavirus vaccines [44]. Similarly, Twitter contents have also been applied to excavate public sentiments and opinions toward COVID-19 vaccines [45]. Similar studies have bolstered the notion that social media can offer valuable illumination to infoveillance, promoting vaccine uptake, and altering vaccine hesitancy. Nevertheless, it should be noted that this series of studies have often been conducted in Western contexts. As a country with an increasingly expanded proportion in social media usage, China has not gained enough scholarly attention.

Based on the aforementioned discussion, this research aims to explore expression differences between professionals and laypeople toward the COVID-19 vaccine on social media. This research topic is essential because it affords a basis for understanding perception disparities between professionals and nonprofessionals, which in turn provides insights into devising effective communication strategies between the 2 groups to promote COVID-19 vaccination compliance and coverage. Additionally, there are limited studies systematically examining expressions between laypeople and professionals [46,47]. Whether the abovementioned risk perception divergence and attitudinal difference reflect in expressions is still unknown. Our research endeavors to replenish the present lacuna by offering empirical evidence on how the 2 groups conceive medical technologies in a public health crisis.

Given China's low visibility in the previous research scope, we focused on China. China was one of the first countries severely affected by COVID-19. After implementing a series of strict prevention and control measures, the Chinese government tamed the virus in a comparatively short period; the so-called "China's model to combat the COVID-19" set an example for other countries to combat this global health crisis [14]. Furthermore, China has taken great strides in developing COVID-19 vaccines. For instance, the 2 Chinese pharmaceutical pioneers Sinovac and Sinopharm have undertaken tremendous vaccine production tasks and promoted their products domestically and overseas [48]. As one of the first-tier countries launching vaccines against COVID-19, the COVID-19 vaccine entered the Chinese public discussion sphere early, endowing us a unique opportunity to unravel the possible asymmetric perceptions between medical professionals and the public toward the same issue. In summary, we formulated two research questions: (1) is there any difference in expression between professionals and laypeople when discussing the COVID-19 vaccine in China? (2) What major themes about the COVID-19 vaccine emerged in the 2 groups' expressions in the Chinese context? Do thematic disparities exist? The first question leans to the explicit layer and focuses on the primary text features. The second question leans to the implicit layer and targets the latent thematic structures. We believe that this study could develop an in-depth understanding of the differences between professionals and laypeople by synthesizing the 2 aspects.

Methods

Data Source

We selected a web-based question and answer (Q&A) forum to collect the research data. Zhihu [49], a Chinese equivalent of Quora, is the most popular social Q&A website in China [46]. According to Liang et al [46], Zhihu is an ideal platform to investigate differences between professionals and laypeople for 3 reasons. First, Zhihu amasses a substantial amount of user-generated content about controversial social issues. For example, as of May 12, 2021, the "COVID-19 vaccine" topic on Zhihu has garnered 762 questions. Second, Zhihu has a unique structure that facilitates interactive communication. Users can follow each other, invite others to answer questions, and reply to each other in the comments section. Third, professionals are highly visible and active on Zhihu. A significant proportion of experts could be easily distinguished by their self-reported personal details (eg, affiliation and working sector) or visual symbols bestowed by the platform (eg, a blue badge after the username) [46]. Those who specialize in particular fields and engage in sharing opinions are more likely to become influencers on Zhihu [50]. These characteristics enable us to discern professionals from laypeople cost-effectively and discover the expression incongruities between the 2 user groups on Zhihu.

To obtain as much comprehensive data as possible, one of the authors designed a Python script to crawl all questions (including extended question descriptions) and their corresponding number of answers under the "COVID-19 vaccine" topic, which is the most relevant and active topic about the COVID-19 vaccines on Zhihu. Since some questions received very few responses, we excluded those questions with less than 10 answers. Next, we adopted another self-written Python script to collect each answer's concrete content along with each respondent's public profile. The content serves as the core corpus of the current study, whereas the public profiles are used to determine the identity category of the respondent. Finally, 65 questions were retained for the ensuing analysis with 3196 answers under them. Multimedia Appendix 1 provides details regarding the reserved questions. Data collection was finished on March 23, 2021.

Coding Scheme

Manual coding was applied to differentiate the 2 types of identities and classify the 65 retained questions. According to the Merriam-Webster Dictionary, a professional can be defined as someone who conforms to the technical or ethical standards of a profession [51]. Because of the inherent medical attributes of COVID-19 vaccines, we further narrowed the meaning scope of *professional* by restricting it to medical professionals. Two criteria were set to distinguish the professional identity: (1) users licensed or certified to provide health care services to natural persons (eg, physicians and pharmacists) [52] and (2) users who major or conduct research in medicine or related fields (eg, Chinese pharmacy or life sciences) [46]. Laypeople are also evaluated on the basis of two criteria: (1) users who explicitly disclose their identities, other than medical professionals and (2) users who do not divulge their identities explicitly. Identification cues are extracted from pertinent information units in the user's public profile, including self-reported educational experience, working sectors, career history, and authentication information.

With regard to the reserved 65 questions, it is untenable to perform between-group comparisons 65 times. In other words, it is not sensible to compare professionals' and laypeople's expressions under each question because it would be difficult to draw a representative and systematic conclusion through repeated small-scale analysis. Therefore, we classified those questions to find out some common underlying characteristics among them. In line with previous experience [53], we carried out semiopen coding to clarify question categories. All authors discussed the classification framework back and forth on the basis of personal understanding after reviewing all questions and their descriptions. Later, we performed a pilot manual coding to confirm the rationality and applicability of the preliminary categories. The final classification comprises 5 categories (Table 1), which suit all questions well. The mapping relationships between individual questions and categories can also be found in Multimedia Appendix 1. More specifically, the 5 categories in Table 1 resonate with preceding studies. Firstly, people's COVID-19 vaccination intention primarily hinges on the safety and side effects of the relevant vaccines [54]. COVID-19 vaccines' efficacy and safety profile are vital for its successful deployment and the achievement of herd immunity [6,9]. Thus, "adverse reactions" and "vaccine effectiveness" are 2 indispensable categories when discussing the COVID-19 vaccine. Secondly, one study about discerning topics regarding vaccines on the internet proposed that disease outbreaks, vaccine development, vaccine studies, and vaccination guidelines emerged in web-based articles on vaccines [55]. Besides, many scholars accentuated vaccines' nonnegligible role in preventing communicable diseases and indicate the severity and hidden threats resulting from vaccine hesitancy from a societal perspective [2,56,57]. Our remaining 3 question categories (Table 1) have significant overlap with those findings.

Table 1. Question categories and their meanings.

Category	Meaning
Adverse reactions	Asking about any unintended or dangerous human reactions to COVID-19 vaccines
Vaccination	Asking about COVID-19 vaccination programs, arrangements, intentions, and status quo
Vaccine effectiveness	Asking about the physiological reactions in individuals, such as the effectiveness and success signs of a specific type of COVID-19 vaccine or efficacy comparison between candidate vaccines
Social implications of the vaccine	Asking about the social consequences of the emergence and uptake of the COVID-19 vaccine, such as whether COVID-19 vaccines can achieve herd immunity
Vaccine development	Asking about details regarding the COVID-19 vaccine development process, such as performance indicators in the 3 trial phases

Analytical Strategies

We selected traditional content analysis and automated text analysis as our research methods to address the 2 proposed research questions. Conventional content analysis aimed to distinguish the identity of each respondent through manual coding. Three authors coded 50 randomly sampled respondents in accordance with the aforementioned designated criteria in the pilot coding stage. Intercoder reliability reached an ideal state (Krippendorff α=.93). The 3 authors then coded the remaining respondents independently. Similarly, 3 authors coded 20 randomly selected questions to test intercoder reliability for the question category. The reliability coefficient also meets the statistical standard (Krippendorff α=.91).

Owing to the large volume of answers, we leveraged automated text analysis to analyze the corpus efficiently. Automated text analysis is a broad terminology for a series of natural language processing methods, including but not limited to frequency analysis, co-occurrence analysis, and topic modeling [58]. This automated approach benefits text miners in alleviating the labor-intensive task of coding texts manually. More specifically, we calculated the fundamental expression characteristics of the 2 user groups, including the answer length, distribution of attitudes, and high-frequency words [46,59]. Attitudinal analysis was completed using the up-to-date TextMind software developed by the Chinese Academy of Science, which can be regarded as the Chinese version of LIWC (Linguistic Inquiry and Word Count) [60]. TextMind is capable of inferring emotional states, intentions, and thinking styles from text through a dictionary-based approach with high reliability and validity [61].

For thematic analysis, we utilized topic modeling to probe into the thematic differences between the 2 identities. Topic modeling can investigate the hidden thematic structure of a given collection of texts [62]. As one of the cutting-edge branches in the topic modeling family, structural topic modeling (STM) allows researchers to estimate a topic model by considering document-level metadata. In other words, STM enables researchers to discover relationships between topics and metadata, such as the topic preference of distinct authors or topic fluctuation across time [63]. STM assimilates document metadata (eg, authorship and time of publication) as covariates during the generative process; it has previously been used to explore the distinct selective sharing mechanisms of different media outlets [64] and how party identification affects topic prevalence [65]. Before formal modeling, the authors conducted preprocessing to clean the corpus, including discarding punctuation, filtering out stop-words, and pruning highly frequent words. The preprocessing procedure adheres to that of a widely recognized topic modeling study [62]. STM was implemented using the stm package in R [63], while other automated text analyses were accomplished in the Python programming environment.

Results

The first research question asks about the expression differences between professionals and laypeople. Given the 5 predefined question categories, we examine all answers under each question category and performed statistical analysis (Tables 2-5).

Compared to the answers of professionals, those of laypeople are more prevalent (Table 2). Besides, professionals are inclined to write longer answers than laypeople (Table 3). A subsequent series of 2-tailed independent-samples *t* tests confirmed this supposition by revealing that professionals' average answer length was significantly higher in word count than that of laypeople under each question category (adverse reactions: t_{711}=–2.335; *P*=.02; vaccination: t_{958}=–2.401; *P*=.02; vaccine effectiveness: t_{415}=–2.240; *P*=.03; social implications of vaccine: t_{260}=–2.149; *P*=.04; vaccine development: t_{842}=–4.546; *P*<.001).

Table 2. The number of answers posted by professionals and laypeople under 5 question categories regarding COVID-19 vaccines (N=3196).

Question category	Answers, n (%)
Adverse reactions	
Professional	68 (9.54)
Laypeople	645 (90.46)
Vaccination	
Professional	104 (10.83)
Laypeople	856 (89.17)
Vaccine effectiveness	
Professional	76 (18.23)
Laypeople	341 (81.77)
Social implications of the vaccine	
Professional	25 (9.54)
Laypeople	237 (90.46)
Vaccine development	
Professional	129 (15.28)
Laypeople	715 (84.72)

Table 3. Answer length of professionals and laypeople under 5 question categories regarding COVID-19 vaccines (N=3196).

Question category	Answer word count, mean (SD)
Adverse reactions	
Professional	454.12 (674.09)
Laypeople	251.83 (806.92)
Vaccination	
Professional	510.67 (1191.63)
Laypeople	225.97 (482.32)
Vaccine effectiveness	
Professional	937.03 (2408.93)
Laypeople	310.80 (619.62)
Social implications of the vaccine	
Professional	765.52 (1310.93)
Laypeople	200.10 (331.42)
Vaccine development	
Professional	815.60 (1345.11)
Laypeople	266.18 (609.15)

Table 4. Attitude distribution of professionals and laypeople 5 five question categories regarding COVID-19 vaccines (N=3196).

Question category	Answers with a positive attitude, n (%)	Answers with a neutral attitude, n (%)	Answers with a negative attitude, n (%)
Adverse reactions			
Professional	21 (30.88)	28 (41.18)	19 (27.94)
Laypeople	209 (32.40)	220 (34.11)	216 (33.49)
Vaccination			
Professional	46 (44.23)	28 (26.92)	30 (28.85)
Laypeople	339 (39.60)	276 (32.24)	241 (28.15)
Vaccine effectiveness			
Professional	38 (50.00)	13 (17.11)	25 (32.89)
Laypeople	170 (49.85)	97 (28.45)	74 (21.70)
Social implications of the vaccine			
Professional	10 (40.00)	6 (24.00)	9 (36.00)
Laypeople	96 (40.51)	67 (28.27)	74 (31.22)
Vaccine development			
Professional	53 (41.09)	49 (37.98)	27 (20.93)
Laypeople	336 (46.99)	219 (30.63)	160 (22.38)

Table 5. High-frequency words of professionals and laypeople under 5 question categories regarding COVID-19 vaccines.

Question category	High-frequency words[a]
Adverse reactions	
Professional	RNA, Pfizer, adverse reactions, death, America, side effects, clinical trial, inject, inactivated vaccine, data
Laypeople	America, China, Pfizer, coronavirus, RNA, death, Japan, inject, adverse reactions, country
Vaccination	
Professional	coronavirus, crowd, immune, infect, clinical trial, antibody, country, adverse reactions, disease, emergency
Laypeople	coronavirus, Russia, America, country, China, crowd, inject, clinical trial, infect, research and development
Vaccine effectiveness	
Professional	RNA, coronavirus, data, protein, effective rate, infect, cell, immune, inactivated vaccine, technology
Laypeople	RNA, China, coronavirus, data, inactivated vaccine, America, India, technology, produce, protein
Social implications of the vaccine	
Professional	coronavirus, data, clinical trial, Sinovac, infect, come into the market, China, symptom, effective rate, country
Laypeople	country, coronavirus, price, China, research and development, control, domestic, America, free of charge, crowd
Vaccine development	
Professional	clinical trial, coronavirus, RNA, experiment, research, research and development, China, infect, clinic, data
Laypeople	America, China, RNA, coronavirus, research and development, country, pregnant woman, experiment, infect, company

[a] The 10 most frequent words are listed, and words are translated from Chinese to English. Some Chinese words correspond to more than 1 English word.

Furthermore, statistical analysis revealed that a positive attitude dominated the discussion regarding COVID-19 vaccines (Table 4). A series of chi-square tests were conducted to examine the correlation between attitude and identity. The results revealed nonsignificant relationships under 4 question categories, which suggests that professionals do not differ significantly from laypeople with respect to their attitude distribution when discussing adverse reactions (χ^2_2=1.5; *P*=.47), vaccination (χ^2_2=1.3; *P*=.51), social implications of the vaccine (χ^2_2=0.3; *P*=.86), and vaccine development (χ^2_2=2.8; *P*=.25). However, for vaccine effectiveness, the correlation reached significance (χ^2_2=6.3; *P*=.04). Post hoc analysis based on the adjusted residual (AD) score revealed that laypeople were less likely to express a negative attitude (AD=–2.100), while professionals favor a negative attitude (AD=2.100) under this category.

With respect to the high-frequency words among the 2 user groups, it is evident that laypeople mentioned countries more frequently (eg, *America*, *China*, *Japan*, *Russia*, and *India*) than professionals. Professionals talked more about medical jargon (eg, *clinical trial*, *immune*, *antibody*, *cell*, and *effective rate*) than laypeople (Table 5). However, a comparison of high-frequency words barely reveals a general word use preference pattern; the latent semantic structures still require a more in-depth inspection. Thus, we performed subsequent STM to deepen our understanding of the 2 groups' topic preferences.

The second research question makes an inquiry about the latent themes that belong to the 2 kinds of identities under the 5 categories and accompanying possible thematic differences. For an accurate and robust estimation, we took advantage of the data-driven approach to select the number of topics, which is a built-in function in the stm package [63]. Based on the semantic coherence and residual fluctuation from multiple rounds of automated tests, we determined the topic number of each question category. The detailed indicators are exhibited in Multimedia Appendix 2.

According to a prior study using STM [13], the topic estimation process sticks to some assumptions. First, each document can be regarded as a mixture of latent topics, where each topic is a probability distribution of words. Second, a document is statistically generated by an iterative inference process. A topic is randomly sampled in each process, and a certain word associated with the topic is randomly drawn. The most probable topics and pertinent distributions are estimated on the basis of the given data. Although the probability distribution of words has no intuitive meaning, researchers can interpret the topic's meaning from the relative importance (or the so-called "weight") of words. In the current study, after executing the STM, topics were represented as collections of words. The authors labeled each topic and summarized the topic's meaning by considering the highest-probability words and exclusive words simultaneously [63]. In STM, words with the highest probabilities and the highest frequency and exclusivity (FREX) weights are provided. A high probability implies that corresponding words are highly likely to appear under the given topic [63], while a high FREX score replenishes the high probability indicator by considering word exclusivity and frequency simultaneously [13]. Topics extracted from answers under each question category were depicted (Figure 1). Detailed topic meanings are shown in Multimedia Appendix 3. Next, we estimated the relationship between user identity and topic prevalence. The stm package illustrates those relationships with forest plots, reflecting the difference in topical prevalence between professionals and laypeople in a more expressive way.

Figures 2-6 delineate the thematic disparities between the 2 user groups under each question category. The horizontal lines represent CIs. If the CIs for each topic overlap with the dotted vertical line (indicates null effect), this implies that at the 95% CI level, professionals and laypeople do not differ from each other in adopting the topic. For the 3 topics under *adverse reactions* (Figure 2), the "safety of Chinese-made vaccine" topic is more likely to be used by laypeople (β=–.032; *P*=.04). For the 4 topics under vaccination (Figure 3), the two topics "vaccination arrangement for priority groups" (β=.044; *P*<.001) and "urgent approval and prioritization of vaccines" (β=.052; *P*<.001) were primarily associated with professionals. In contrast, the other 2 topics "vaccines in Russia" (β=–.037; *P*<.001) and "the effectiveness of vaccination in Russia and the U.S." (β=–.059; *P*<.001) were more frequently adopted by laypeople. Among the 3 topics under *vaccine effectiveness* (Figure 4), 2 varied significantly across the 2 user groups. "indicators for evaluating vaccine effectiveness" topic (β=–.044; *P*=.003) was more likely to be mentioned by laypeople, while "medical principles of vaccine effectiveness" (β=.026; *P*=.03) was more inclined to be mentioned by professionals. Regarding the 4 topics under *social implications of the vaccine* (Figure 5), none of them reached significantly difference levels. Regarding the last category (Figure 6), "principles of vaccine trials" (β=.139; *P*<.001) was more inclined to be mentioned by professionals. Conversely, "vaccine development process worldwide" (β=–.132; *P*<.001) was more inclined to be mentioned by laypeople.

Figure 1. Question categories and their related topics under the COVID-19 vaccine issue on Zhihu.

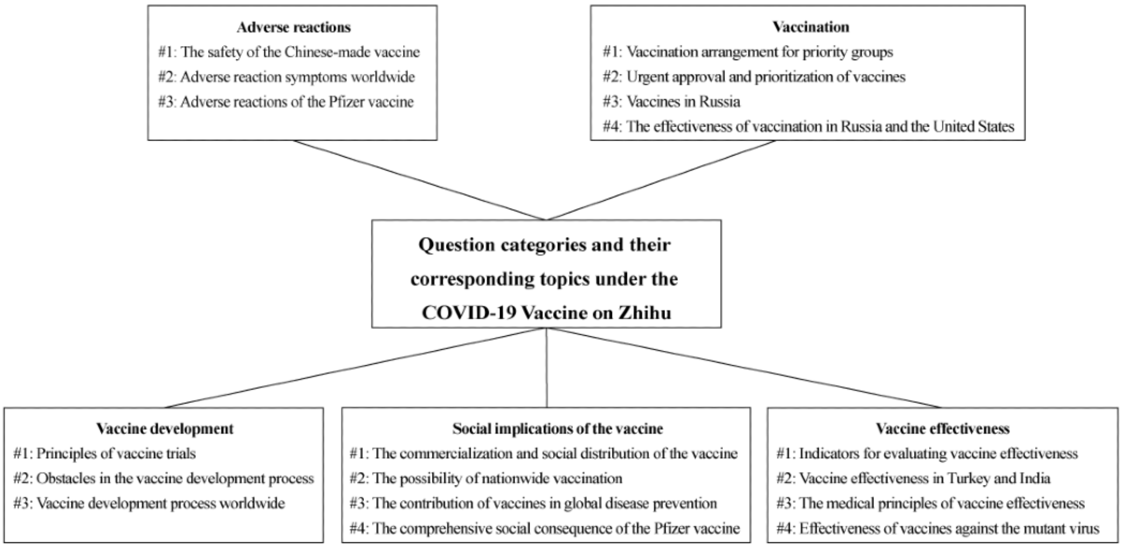

Adverse reactions
#1: The safety of the Chinese-made vaccine
#2: Adverse reaction symptoms worldwide
#3: Adverse reactions of the Pfizer vaccine
Vaccination
#1: Vaccination arrangement for priority groups
#2: Urgent approval and prioritization of vaccines
#3: Vaccines in Russia
#4: The effectiveness of vaccination in Russia and the United States
Question categories and their corresponding topics under the COVID-19 Vaccine on Zhihu
Vaccine development
#1: Principles of vaccine trials
#2: Obstacles in the vaccine development process
#3: Vaccine development process worldwide
Social implications of the vaccine
#1: The commercialization and social distribution of the vaccine
#2: The possibility of nationwide vaccination
#3: The contribution of vaccines in global disease prevention
#4: The comprehensive social consequence of the Pfizer vaccine
Vaccine effectiveness
#1: Indicators for evaluating vaccine effectiveness
#2: Vaccine effectiveness in Turkey and India
#3: The medical principles of vaccine effectiveness
#4: Effectiveness of vaccines against the mutant virus

Figure 2. Thematic disparities between professionals and laypeople under the "adverse reactions" question category.

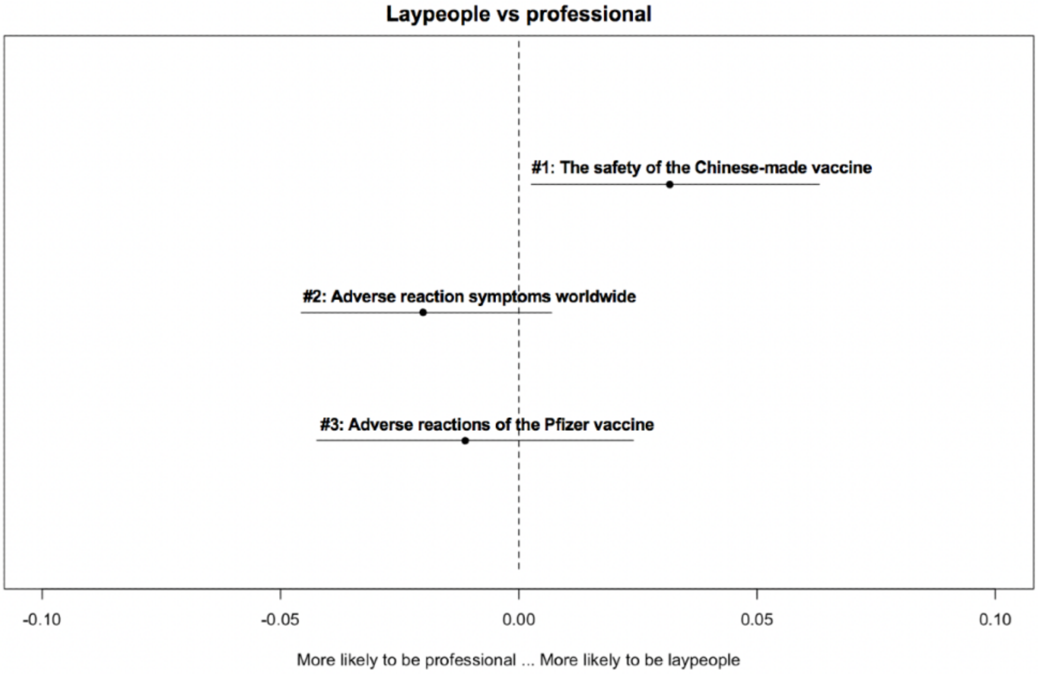


Figure 3. Thematic disparities between professionals and laypeople under the "vaccination" question category.

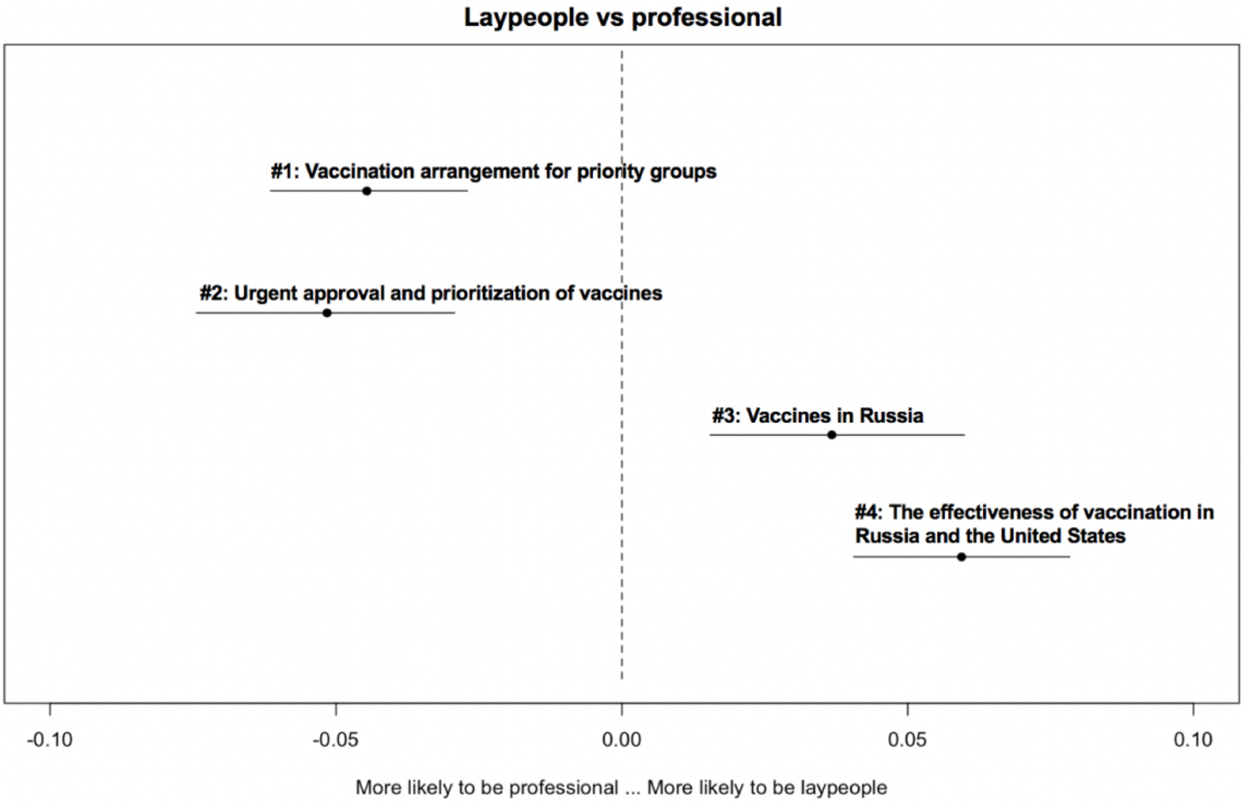


Figure 4. Thematic disparities between professionals and laypeople under the "vaccine effectiveness" question category.

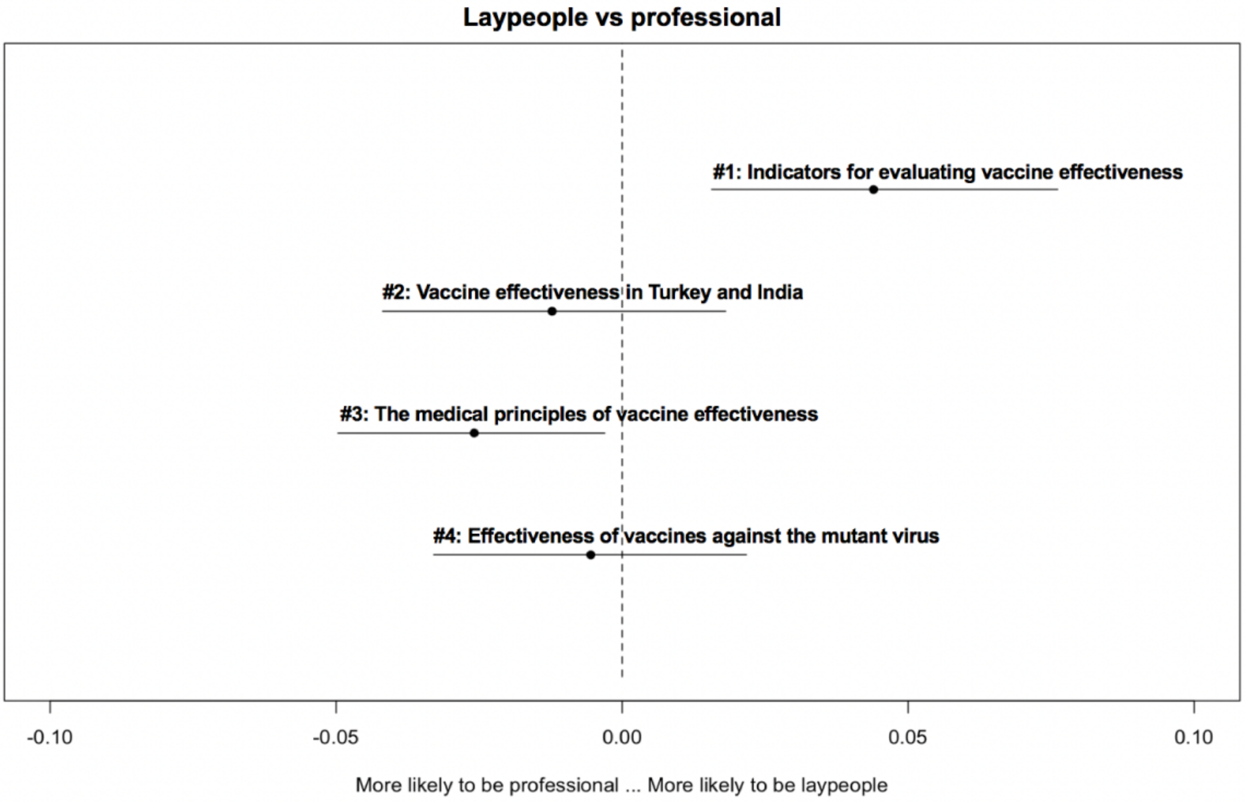


Figure 5. Thematic disparities between professionals and laypeople under the "social implications of the vaccine" question category.

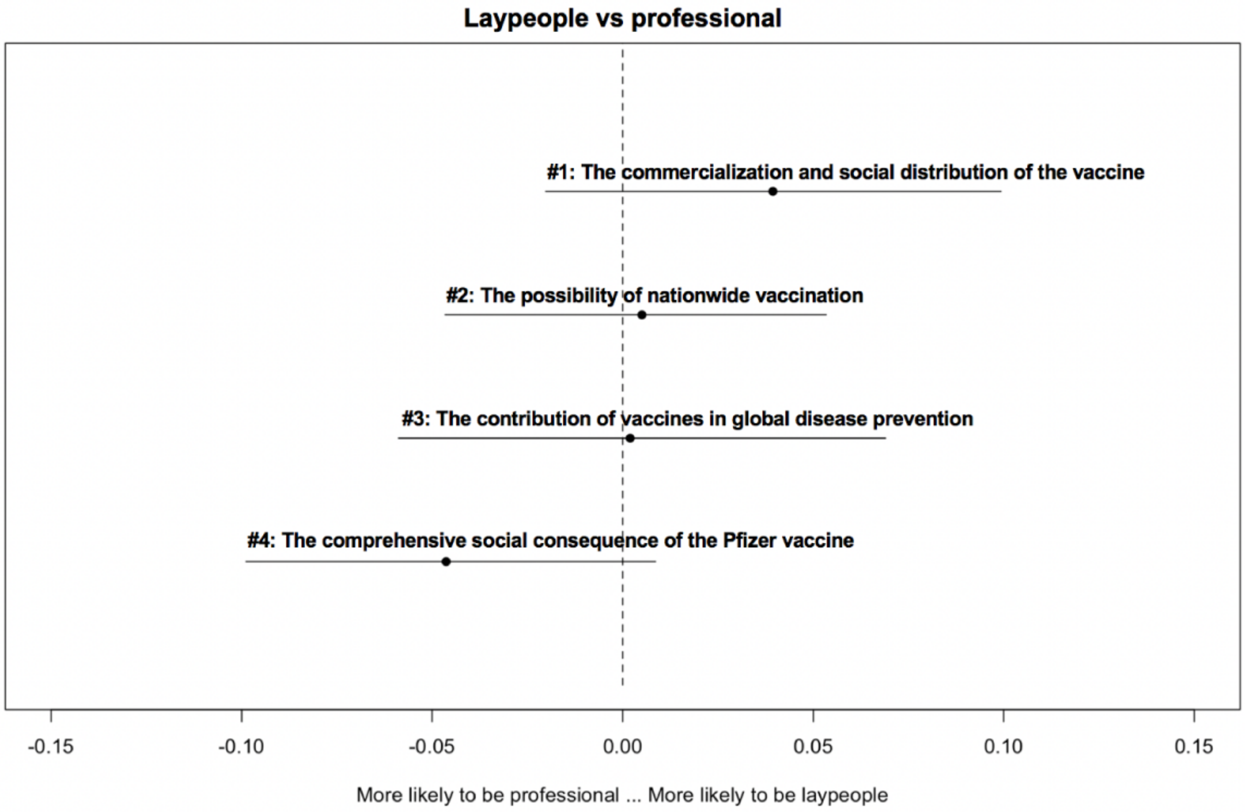
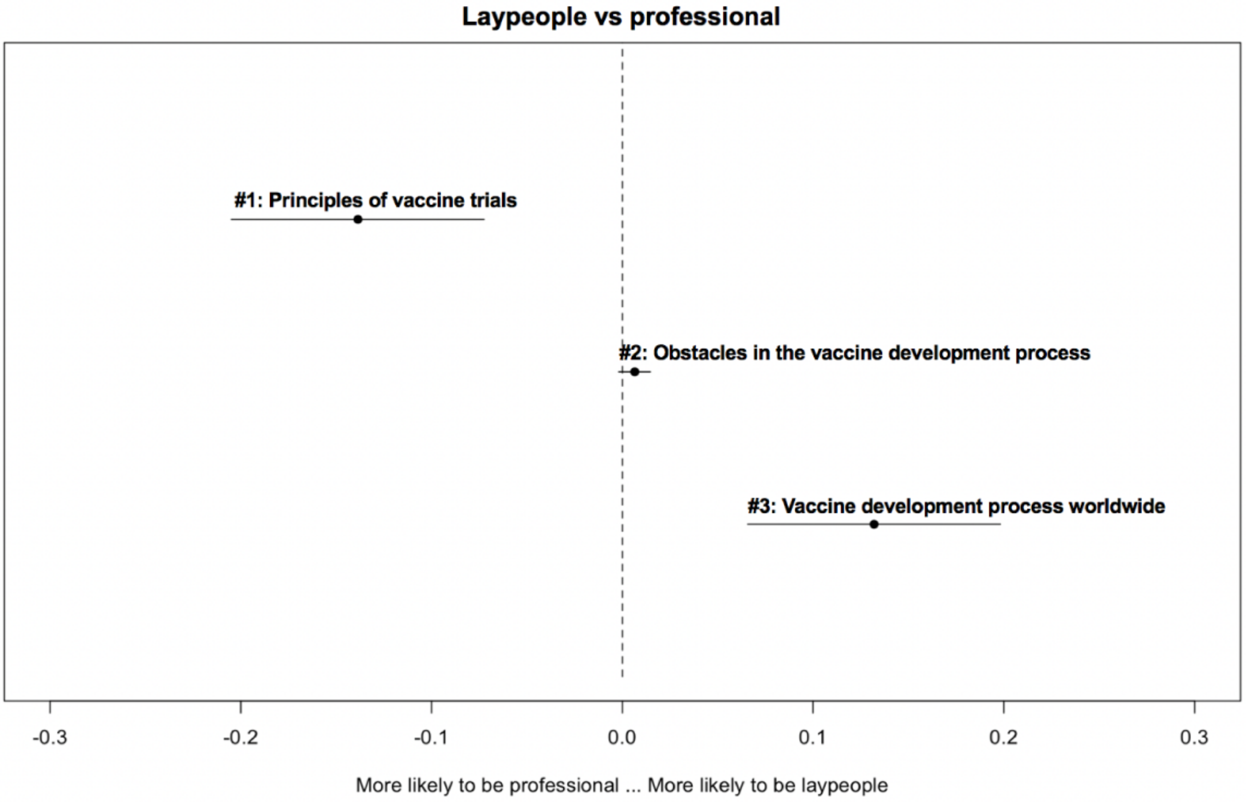


Figure 6. Thematic disparities between professionals and laypeople under the "vaccine development" question category.



Discussion

Principal Findings

This study aimed to disentangle the expression differences between professionals and laypeople in the context of a somewhat contentious issue. To the best of our knowledge, this is one of the few studies adopting STM to analyze thematic disparities between these 2 user groups, which goes beyond previous studies that mainly relied on the hand-annotated method [46]. Moreover, there is a shortage of studies focusing on the professional-laypeople divide during the COVID-19 pandemic. Our study contributes to comprehending the expression characteristics of the 2 identities and provides us an empirical foundation for facilitating professional-laypeople communication in a web-based Q&A environment, further helps advocate authoritative voices, and corrects misinformation in a time inundated with uncertainties and risks [66].

Per our primary findings, the first arresting finding is the active participation of laypeople in the COVID-19 vaccine issue. This phenomenon, to some extent, gives credence to the previous viewpoint on the communication-facilitating effect of social media. Brossard [67] contended that the new media technologies afford the lay audience more opportunities to participate in and discuss scientific issues in a relatively straightforward way. Similarly, Peters [68] bolsters this assertion by reporting that circumstances for web-based communication substantially challenge the once quasi-monopoly status of intermediary information disseminators (eg, professional journalists and scientists) [68]. Therefore, although laypeople do not possess equivalent professional knowledge as professionals, the former are still guaranteed sufficient opportunities to discuss professional issues with professionals. In other words, the social media platforms characterize equality, openness, and plurality, which lowers the knowledge threshold and entry barrier when discussing medical issues. However, whether this frequent occurrence of laypeople equates to effective communication or fruitful dialogue between these 2 groups needs further investigation.

Aside from the extensive participation of laypeople, our study revealed additional expression differences between the 2 user groups. First, the average answer length of professionals was longer than that of laypeople. Backed with professional knowledge and practical experience, professionals are likely to elaborate their viewpoints by incorporating various evidence. This is especially true for the COVID-19 vaccine topic because COVID-19 is a typical “sudden and unexpected event” [69] with medical puzzles, and the COVID-19 vaccine still calls for rigorous clinical trials and continuous surveillance [4]. According to Zou et al [70], statistical evidence and narrative evidence are 2 major types of evidence adopted to elucidate health-related topics. Professionals are more familiar with quantitative and numerical evidence owing to their professional background and working experience. They can also invoke narrative evidence derived from daily experiences to support their views. However, laypeople lack quantitative arguments and have to depend on narratives to expound their viewpoints. Furthermore, professionals may have a more cautious and conservative mindset because of the intrinsic features in their vocational training and educational background. One representative example is professionals are not as optimistic as laypeople when talking about vaccine effectiveness on the premise that COVID-19 vaccine development is an ongoing process that requires more reliable evidence, such as the undetermined age-specific adverse effects [71].

Our results also show that professionals and laypeople analyzed the COVID-19 vaccine issue from varying perspectives. Echoing the literature review, 1 long-standing speculation in the public health field and science communication fields is that laypeople’s risk perceptions are always insufficient with regard to scientific assessments [72]. The scientific knowledge deficiency among the lay public hampers their ability to understand specific scientific issues and establish a positive attitude toward them [38,39,73]. Considering risk perception and attitude together, we prefer to believe that laypeople’s knowledge is not quantitatively lesser than or qualitatively inferior to that of professionals. Instead, the 2 user groups share some similarities but hold different thinking angles simultaneously, which is more appropriate to be marked as “qualitatively different.” First, the 2 user groups unanimously paid attention to adverse reaction symptoms worldwide, the vaccine’s effectiveness against the mutant virus, the contribution of vaccination for global disease prevention, and some other topics, which implies overlaps in their perspectives. However, considering issues related to medical expertise, such as the vaccination question category in our study, professionals accentuate arrangement and urgent approval, which are inextricably linked to public policies, and the reasonable allocation of medical resources. Laypeople prefer to care about other countries, presumably driven by the overwhelming media coverage on epidemic situations in other countries. This comparison suggests that the disparities rest in the division between professional and experiential modes of thinking, which act as 2 thinking modes toward controversial issues. The stark contrast also manifests in high-frequency word comparison and other medical-related question categories, including vaccine development and effectiveness. Second, we did not observe clear distinctions between the 2 user groups with regard to attitude under 4 question categories, which further illustrates that the attitudinal difference assumption based on knowledge level disparities is untenable in the Chinese COVID-19 vaccine context. Despite some objective gaps in knowledge acquisition between professionals and laypeople, they were both willing to treat the COVID-19 vaccines positively. Third, the “adverse reactions” category is most closely related to risk. In fact, we did not see laypeople lay excessive stress on the abnormal symptoms. This finding debunks the risk perception disparities that originated from the knowledge deficiency supposition, which implies that laypeople are not always amplifying the risks. They favor countries’ specific situations and think from living experience rather than magnifying vaccine risks or expressing suspicion regarding COVID-19 vaccines.

Regarding the social implications of the vaccines, as a category not closely linked to medical knowledge, the 2 user groups showed no significant differences. This finding indicates that the professional and experimental thinking modes lost their explanatory power when encountering the abstract issue. The social implications of COVID-19 vaccines can be broad and intricate, related to a wide range of societal dimensions. Hence, it is difficult for professionals or laypeople to lay particular emphasis on merely 1 mode. Combining the topics' similarities and incongruities between the 2 user groups, we conclude that apart from the overlaps, the "qualitatively different" characteristic is also common on the web-based Q&A forum, which reflects different perspectives derived from knowledge background and life experience. In the context of COVID-19 vaccines, the medical-related questions are more sensitive to the influence of the "qualitatively different" feature, while more broad and abstract questions seem impervious to this feature.

Limitations

Our analysis bears several caveats. With respect to the question categories, the COVID-19 vaccine is a multifaceted, intricate, and context-dependent issue associated with copious aspects [5]. Some question categories, such as vaccines and international relations, are omitted in this study and hence need to be further explored in future studies. Besides, the inclusion of longitudinal perspectives in this text mining study would yield more intriguing findings. For instance, with the development of the COVID-19 pandemic, will the thematic differences between these 2 user groups become wider or narrower? A dynamic and longitudinal approach would undoubtedly advance our comprehension of the ongoing COVID-19 vaccine issue and help curb this public health emergency. Furthermore, 1 aspect that cannot be dismissed is that the answers, of both professionals and laypeople, were largely hinged on the characteristics of the questions. Thus, the topic distribution may be confined within the questions' scopes. Future studies could focus on other social media platforms (eg, Twitter and Sina Weibo) to obtain a more holistic discursive landscape, which may be more topic-rich owing to the absence of designated questions.

Conclusions

This study provides an overview of opinion patterns and scrutinizes the expression differences between professionals and laypeople toward the COVID-19 vaccine. In terms of quantity, laypeople are the dominant discussants in the web-based Q&A forum Zhihu. Regarding expression differences, the professionals preferred writing longer answers than laypeople; they also showed a conservative stance in vaccine effectiveness and tended to mention medical terminologies in their discussions. By exerting the power of STM, as a valuable tool under unsupervised machine learning, we outlined the topics under each question category, along with the topic preference of the 2 groups. In a nutshell, professionals paid more attention to the medical principles and professional standards nested in discourses on COVID-19 vaccines. In contrast, laypeople showed solicitude explicitly for vaccine-related issues at the national and global levels, and to the safety of the Chinese-made vaccine. The 2 user groups shared some common grounds and manifested distinct concerns within the COVID-19 vaccine context.

We believe that this study has some implications and merits. First, public health scholars should be keenly aware of expressions and discussions on web-based Q&A forums, which were comparatively overlooked in prior infoveillance or infodemiology studies [74]. Q&A forums such as Zhihu or Quora make a clear distinction between professionals and laypeople, thus providing researchers with opportunities to explore the professional-laypeople incongruities in discursive styles and core topics. These dimensions may further facilitate addressing the underlying "distance" or "gap" between the 2 user groups [68]. Second, extant studies germane to COVID-19–related topic modeling widely to probe into public concerns and public awareness [75,76]. However, there is a paucity of studies on the thematic differences among various identities. Our attempts using STM provide a viable solution to discover the nuanced differences between distinct identities, unfolding some particular advantages over traditional topic modeling. Third, for public health educators, effective professional-laypeople communication does not need to focus on all underlying topics. Considering the "qualitatively different" characteristic, practitioners should focus on discussing topics that are significantly inconsistent across different identities and strive to mitigate misunderstanding while generating consensus on those topics. For example, some scholars found that popular conspiracies on Chinese social media, which are related to the pandemic's origin, are about whether country actors intentionally developed SARS-CoV-2 in the laboratory or as bioweapons [77]. Since laypeople are highly concerned with COVID-19 vaccines in foreign countries, public health practitioners must closely scrutinize relevant discussions to guard against the emergence of vaccine-related rumors, conspiracies, or hate speech and strive to create an atmosphere for a rational discussion.

Acknowledgments

The authors wish to thank Dr Anfan Chen (postdoctoral researcher, The Chinese University of Hong Kong) for his constructive suggestions in the initial stage of the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Question categories and corresponding questions with their number of answers.
[DOC File , 80 KB - jmir_v23i8e30715_app1.doc]

Multimedia Appendix 2
Using semantic coherence and residual fluctuation to determine the number of topics (k) under each question category.
[DOC File , 706 KB - jmir_v23i8e30715_app2.doc]

Multimedia Appendix 3
Topics, topic meaning, and corresponding keywords under each question category.
[DOC File , 41 KB - jmir_v23i8e30715_app3.doc]

References

1. WHO Coronavirus (COVID-19) Dashboard. World Health Organization. URL: https://covid19.who.int/ [accessed 2021-04-21]
2. Mo PK, Luo S, Wang S, Zhao J, Zhang G, Li L, et al. Intention to Receive the COVID-19 Vaccination in China: Application of the Diffusion of Innovations Theory and the Moderating Role of Openness to Experience. Vaccines (Basel) 2021 Feb 05;9(2):129 [FREE Full text] [doi: 10.3390/vaccines9020129] [Medline: 33562894]
3. Kaur SP, Gupta V. COVID-19 Vaccine: A comprehensive status report. Virus Res 2020 Oct 15;288:198114 [FREE Full text] [doi: 10.1016/j.virusres.2020.198114] [Medline: 32800805]
4. Haynes BF, Corey L, Fernandes P, Gilbert PB, Hotez PJ, Rao S, et al. Prospects for a safe COVID-19 vaccine. Sci Transl Med 2020 Nov 04;12(568):eabe0948. [doi: 10.1126/scitranslmed.abe0948] [Medline: 33077678]
5. Lazarus JV, Ratzan SC, Palayew A, Gostin LO, Larson HJ, Rabin K, et al. A global survey of potential acceptance of a COVID-19 vaccine. Nat Med 2021 Feb;27(2):225-228 [FREE Full text] [doi: 10.1038/s41591-020-1124-9] [Medline: 33082575]
6. Ruiz JB, Bell RA. Predictors of intention to vaccinate against COVID-19: Results of a nationwide survey. Vaccine 2021 Feb 12;39(7):1080-1086 [FREE Full text] [doi: 10.1016/j.vaccine.2021.01.010] [Medline: 33461833]
7. Malik AA, McFadden SM, Elharake J, Omer SB. Determinants of COVID-19 vaccine acceptance in the US. EClinicalMedicine 2020 Sep;26:100495 [FREE Full text] [doi: 10.1016/j.eclinm.2020.100495] [Medline: 32838242]
8. Murphy J, Vallières F, Bentall RP, Shevlin M, McBride O, Hartman TK, et al. Psychological characteristics associated with COVID-19 vaccine hesitancy and resistance in Ireland and the United Kingdom. Nat Commun 2021 Jan 04;12(1):29 [FREE Full text] [doi: 10.1038/s41467-020-20226-9] [Medline: 33397962]
9. Haque A, Pant AB. Efforts at COVID-19 Vaccine Development: Challenges and Successes. Vaccines (Basel) 2020 Dec 06;8(4):739 [FREE Full text] [doi: 10.3390/vaccines8040739] [Medline: 33291245]
10. Eibensteiner F, Ritschl V, Nawaz FA, Fazel SS, Tsagkaris C, Kulnik ST, et al. People's Willingness to Vaccinate Against COVID-19 Despite Their Safety Concerns: Twitter Poll Analysis. J Med Internet Res 2021 Apr 29;23(4):e28973 [FREE Full text] [doi: 10.2196/28973] [Medline: 33872185]
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. Finney Rutten LJ, Blake KD, Greenberg-Worisek AJ, Allen SV, Moser RP, Hesse BW. Online Health Information Seeking Among US Adults: Measuring Progress Toward a Healthy People 2020 Objective. Public Health Rep 2019;134(6):617-625 [FREE Full text] [doi: 10.1177/0033354919874074] [Medline: 31513756]
13. 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 Jun 02;22(6):e19455 [FREE Full text] [doi: 10.2196/19455] [Medline: 32463367]
14. Luo C, Li Y, Chen A, Tang Y. What triggers online help-seeking retransmission during the COVID-19 period? Empirical evidence from Chinese social media. PLoS One 2020;15(11):e0241465 [FREE Full text] [doi: 10.1371/journal.pone.0241465] [Medline: 33141860]
15. Shen C, Chen A, Luo C, Zhang J, Feng B, Liao W. Using Reports of Symptoms and Diagnoses on Social Media to Predict COVID-19 Case Counts in Mainland China: Observational Infoveillance Study. J Med Internet Res 2020 May 28;22(5):e19421 [FREE Full text] [doi: 10.2196/19421] [Medline: 32452804]
16. Rains SA. Big Data, Computational Social Science, and Health Communication: A Review and Agenda for Advancing Theory. Health Commun 2020;35(1):26-34. [doi: 10.1080/10410236.2018.1536955] [Medline: 30351198]
17. Featherstone JD, Ruiz JB, Barnett GA, Millam BJ. Exploring childhood vaccination themes and public opinions on Twitter: A semantic network analysis. Telematics and Informatics 2020 Nov;54:101474. [doi: 10.1016/j.tele.2020.101474]
18. Featherstone JD, Zhang J. Feeling angry: the effects of vaccine misinformation and refutational messages on negative emotions and vaccination attitude. J Health Commun 2020 Sep 01;25(9):692-702. [doi: 10.1080/10810730.2020.1838671] [Medline: 33103600]
19. Grajales FJ, Sheps S, Ho K, Novak-Lauscher H, Eysenbach G. Social media: a review and tutorial of applications in medicine and health care. J Med Internet Res 2014 Feb 11;16(2):e13 [FREE Full text] [doi: 10.2196/jmir.2912] [Medline: 24518354]
20. Chen S, Xu Q, Buchenberger J, Bagavathi A, Fair G, Shaikh S, et al. Dynamics of Health Agency Response and Public Engagement in Public Health Emergency: A Case Study of CDC Tweeting Patterns During the 2016 Zika Epidemic. JMIR Public Health Surveill 2018 Nov 22;4(4):e10827 [FREE Full text] [doi: 10.2196/10827] [Medline: 30467106]

21. Lwin MO, Lu J, Sheldenkar A, Schulz PJ. Strategic Uses of Facebook in Zika Outbreak Communication: Implications for the Crisis and Emergency Risk Communication Model. Int J Environ Res Public Health 2018 Sep 10;15(9):1974 [FREE Full text] [doi: 10.3390/ijerph15091974] [Medline: 30201929]
22. Sutton J, Renshaw SL, Butts CT. COVID-19: Retransmission of official communications in an emerging pandemic. PLoS One 2020;15(9):e0238491 [FREE Full text] [doi: 10.1371/journal.pone.0238491] [Medline: 32936804]
23. Li Z, Xu X. Analysis of Network Structure and Doctor Behaviors in E-Health Communities from a Social-Capital Perspective. Int J Environ Res Public Health 2020 Feb 11;17(4):1136 [FREE Full text] [doi: 10.3390/ijerph17041136] [Medline: 32053913]
24. Larrouy-Maestri P, Magis D, Grabenhorst M, Morsomme D. Layman versus Professional Musician: Who Makes the Better Judge? PLoS One 2015;10(8):e0135394 [FREE Full text] [doi: 10.1371/journal.pone.0135394] [Medline: 26308213]
25. Feleszko W, Lewulis P, Czarnecki A, Waszkiewicz P. Flattening the Curve of COVID-19 Vaccine Rejection-An International Overview. Vaccines (Basel) 2021 Jan 13;9(1):44 [FREE Full text] [doi: 10.3390/vaccines9010044] [Medline: 33451104]
26. Slovic P. Understanding Perceived Risk: 1978–2015. Environ Sci Policy 2015 Dec 31;58(1):25-29. [doi: 10.1080/00139157.2016.1112169]
27. Slovic P. Trust, emotion, sex, politics, and science: surveying the risk-assessment battlefield. Risk Anal 1999 Aug;19(4):689-701. [doi: 10.1023/a:1007041821623] [Medline: 10765431]
28. DuPont RL. Nuclear Phobia--Phobic Thinking about Nuclear Power. Washington, DC: The Media Institute; 1980.
29. Covello V, Flamm W, Rodricks J, Tardiff R. The analysis of actual versus perceived risks. Boston, MA: Springer; 1983.
30. Slovic P. Risk Perception. In: Travis CC, editor. Carcinogen Risk Assessment. Boston, MA: Springer; 1988:171-181.
31. Kasper RG. Perceptions of Risk and Their Effects on Decision Making. In: Societal Risk Assessment. Berlin: Springer Science & Business Media; 1980:71-80.
32. Bouchez M, Ward JK, Bocquier A, Benamouzig D, Peretti-Watel P, Seror V, et al. Physicians' decision processes about the HPV vaccine: A qualitative study. Vaccine 2021 Jan 15;39(3):521-528. [doi: 10.1016/j.vaccine.2020.12.019] [Medline: 33328141]
33. Chen X, Wagner AL, Zheng X, Xie J, Boulton ML, Chen K, et al. Hepatitis E vaccine in China: Public health professional perspectives on vaccine promotion and strategies for control. Vaccine 2019 Oct 08;37(43):6566-6572. [doi: 10.1016/j.vaccine.2019.07.067] [Medline: 31353258]
34. Alshammari TM, AlFehaid LS, AlFraih JK, Aljadhey HS. Health care professionals' awareness of, knowledge about and attitude to influenza vaccination. Vaccine 2014 Oct 14;32(45):5957-5961. [doi: 10.1016/j.vaccine.2014.08.061] [Medline: 25218193]
35. Stoutenborough JW, Sturgess SG, Vedlitz A. Knowledge, risk, and policy support: Public perceptions of nuclear power. Energy Policy 2013 Nov;62:176-184. [doi: 10.1016/j.enpol.2013.06.098]
36. Ho SS, Xiong R, Chuah AS. Heuristic cues as perceptual filters: Factors influencing public support for nuclear research reactor in Singapore. Energy Policy 2021 Mar;150:112111. [doi: 10.1016/j.enpol.2020.112111]
37. Hansen J, Holm L, Frewer L, Robinson P, Sandøe P. Beyond the knowledge deficit: recent research into lay and expert attitudes to food risks. Appetite 2003 Oct;41(2):111-121. [doi: 10.1016/s0195-6663(03)00079-5] [Medline: 14550309]
38. Miller J, Pardo R, Niwa F. Public Perceptions of Science and Technology: A Comparative Study of the European Union, the United States, Japan, and Canada. Bilbao: Fundación BBV; 1997.
39. Miller J, Kimmel L. Biomedical Communications: Purpose, Audience, and Strategies. Amsterdam: Elsevier; 2001.
40. Ho SS, Leong AD, Looi J, Chen L, Pang N, Tandoc E. Science Literacy or Value Predisposition? A Meta-Analysis of Factors Predicting Public Perceptions of Benefits, Risks, and Acceptance of Nuclear Energy. Environ Commun 2018 Jan 03;13(4):457-471. [doi: 10.1080/17524032.2017.1394891]
41. Leibowitz MK, Scudder MR, McCabe M, Chan JL, Klein MR, Trueger NS, et al. Emergency Medicine Influencers' Twitter Use During the COVID-19 Pandemic: A Mixed-methods Analysis. West J Emerg Med 2021 Mar 22;22(3):710-718 [FREE Full text] [doi: 10.5811/westjem.2020.12.49213] [Medline: 34125051]
42. Zhang J, Xue H, Calabrese C, Chen H, Dang JHT. Understanding Human Papillomavirus Vaccine Promotions and Hesitancy in Northern California Through Examining Public Facebook Pages and Groups. Front. Digit. Health 2021 Jun 17;3:683090. [doi: 10.3389/fdgth.2021.683090]
43. van Atteveldt W, Peng T. When Communication Meets Computation: Opportunities, Challenges, and Pitfalls in Computational Communication Science. Commun Methods Meas 2018 Apr 20;12(2-3):81-92. [doi: 10.1080/19312458.2018.1458084]
44. Surian D, Nguyen DQ, Kennedy G, Johnson M, Coiera E, Dunn AG. Characterizing Twitter Discussions About HPV Vaccines Using Topic Modeling and Community Detection. J Med Internet Res 2016 Aug 29;18(8):e232 [FREE Full text] [doi: 10.2196/jmir.6045] [Medline: 27573910]
45. Yousefinaghani S, Dara R, Mubareka S, Papadopoulos A, Sharif S. An analysis of COVID-19 vaccine sentiments and opinions on Twitter. Int J Infect Dis 2021 Jul;108:256-262 [FREE Full text] [doi: 10.1016/j.ijid.2021.05.059] [Medline: 34052407]
46. Liang J, Liu X, Zhang W. Scientists vs laypeople: How genetically modified food is discussed on a Chinese Q&A website. Public Underst Sci 2019 Nov;28(8):991-1004. [doi: 10.1177/0963662519876782] [Medline: 31547775]

47. Fischhoff B, Scheufele DA. The science of science communication. Introduction. Proc Natl Acad Sci USA 2013 Aug 20;110 Suppl 3(Supplement_3):14031-14032 [FREE Full text] [doi: 10.1073/pnas.1312080110] [Medline: 23942127]
48. Tan Y. Covid: What do we know about China's coronavirus vaccines? BBC News. 2021 Jan 14. URL: https://www.bbc.com/news/world-asia-china-55212787 [accessed 2021-04-21]
49. Zhihu. URL: https://www.zhihu.com/ [accessed 2020-05-25]
50. Zhang J, Chu W. CNKI. URL: https://www.cnki.com.cn/Article/CJFDTotal-SGLK202005006.htm [accessed 2020-06-19]
51. Professional. Merriam-Webster. URL: https://www.merriam-webster.com/dictionary/professional [accessed 2021-04-21]
52. Virginia's Legislative Information System. URL: https://lis.virginia.gov/cgi-bin/legp604.exe?091+ful+SB318&091+ful+SB318 [accessed 2021-04-21]
53. Chen Z, Su CC, Chen A. Top-down or Bottom-up? A Network Agenda-setting Study of Chinese Nationalism on Social Media. J Broadcast Electron Media 2019 Sep 20;63(3):512-533. [doi: 10.1080/08838151.2019.1653104]
54. Liu R, Zhang Y, Nicholas S, Leng A, Maitland E, Wang J. COVID-19 Vaccination Willingness among Chinese Adults under the Free Vaccination Policy. Vaccines (Basel) 2021 Mar 21;9(3):292 [FREE Full text] [doi: 10.3390/vaccines9030292] [Medline: 33801136]
55. Xu Z. Personal stories matter: topic evolution and popularity among pro- and anti-vaccine online articles. J Comput Soc Sc 2019 Apr 9;2(2):207-220. [doi: 10.1007/s42001-019-00044-w]
56. Habibabadi S, Haghighi P. Topic Modelling for Identification of Vaccine Reactions in Twitter. 2019 Presented at: ACSW 2019: Australasian Computer Science Week 2019; January 29-31, 2019; Sydney, NSW. [doi: 10.1145/3290688.3290735]
57. Skeppstedt M, Kerren A, Stede M. Vaccine Hesitancy in Discussion Forums: Computer-Assisted Argument Mining with Topic Models. Stud Health Technol Inform 2018;247:366-370. [Medline: 29677984]
58. Hilbert M, Barnett G, Blumenstock J, Contractor N, Diesner J, Frey S. Computational communication science: A methodological catalyzer for a maturing discipline. Int J Commun 2019;13:3912-3934 [FREE Full text]
59. Li Y, Luo C, Chen A. The evolution of online discussions about GMOs in China over the past decade: Changes, causes and characteristics. Cult Sci 2019 Dec 01;2(4):311-325. [doi: 10.1177/209660831900200406]
60. Song Y, Kwon K, Lu Y, Fan Y, Li B. The “Parallel Pandemic” in the Context of China: The Spread of Rumors and Rumor-Corrections During COVID-19 in Chinese Social Media. Am Behav Sci 2021 Mar 24:000276422110031 [FREE Full text] [doi: 10.1177/00027642211003153]
61. Gao R, Hao B, Li H, Gao Y, Zhu T. Developing Simplified Chinese Psychological Linguistic Analysis Dictionary for Microblog. 2013 Presented at: International Conference on Brain and Health Informatics; October 29-31, 2013; Maebashi. [doi: 10.1007/978-3-319-02753-1_36]
62. Maier D, Waldherr A, Miltner P, Wiedemann G, Niekler A, Keinert A, et al. Applying LDA Topic Modeling in Communication Research: Toward a Valid and Reliable Methodology. Commun Methods Meas 2018 Feb 16;12(2-3):93-118. [doi: 10.1080/19312458.2018.1430754]
63. Roberts ME, Stewart BM, Tingley D. stm: An R Package for Structural Topic Models. J Stat Soft 2019;91(2). [doi: 10.18637/jss.v091.i02]
64. Pak C. News Organizations’ Selective Link Sharing as Gatekeeping: A Structural Topic Model Approach. Comput Commun Res 2019;1(1) [FREE Full text] [doi: 10.31235/osf.io/pt7es]
65. Roberts ME, Stewart BM, Tingley D, Lucas C, Leder-Luis J, Gadarian SK, et al. Structural Topic Models for Open-Ended Survey Responses. Am J Pol Sci 2014 Mar 06;58(4):1064-1082. [doi: 10.1111/ajps.12103]
66. Bode L, Vraga E. Correction Experiences on Social Media During COVID-19. Social Media + Society 2021 Apr 12;7(2):205630512110088 [FREE Full text] [doi: 10.1177/20563051211008829]
67. Brossard D. New media landscapes and the science information consumer. Proc Natl Acad Sci USA 2013 Aug 20;110 Suppl 3:14096-14101 [FREE Full text] [doi: 10.1073/pnas.1212744110] [Medline: 23940316]
68. Peters HP. Gap between science and media revisited: scientists as public communicators. Proc Natl Acad Sci USA 2013 Aug 20;110 Suppl 3:14102-14109 [FREE Full text] [doi: 10.1073/pnas.1212745110] [Medline: 23940312]
69. Lu Y, Pan J, Xu Y. Public Sentiment on Chinese Social Media during the Emergence of COVID19. JQD 2021 Apr 26;1 [FREE Full text] [doi: 10.51685/jqd.2021.013]
70. Zou W, Zhang WJ, Tang L. What Do Social Media Influencers Say about Health? A Theory-Driven Content Analysis of Top Ten Health Influencers' Posts on Sina Weibo. J Health Commun 2021 Jan 02;26(1):1-11. [doi: 10.1080/10810730.2020.1865486] [Medline: 33372857]
71. Lipsitch M, Dean NE. Understanding COVID-19 vaccine efficacy. Science 2020 Nov 13;370(6518):763-765. [doi: 10.1126/science.abe5938] [Medline: 33087460]
72. McComas KA. Defining moments in risk communication research: 1996-2005. J Health Commun 2006;11(1):75-91. [doi: 10.1080/10810730500461091] [Medline: 16546920]
73. Bucchi M, Trench B. Science Communication and Science in Society: A Conceptual Review in Ten Keywords. TECNOSCIENZA Ital J Sci Technol Stud 2016;7(2):151-168 [FREE Full text]
74. 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. [doi: 10.2196/jmir.1157] [Medline: 19329408]

75. Boon-Itt S, Skunkan Y. Public Perception of the COVID-19 Pandemic on Twitter: Sentiment Analysis and Topic Modeling Study. JMIR Public Health Surveill 2020 Nov 11;6(4):e21978 [FREE Full text] [doi: 10.2196/21978] [Medline: 33108310]
76. Xue J, Chen J, Hu R, Chen C, Zheng C, Su Y, et al. Twitter Discussions and Emotions About the COVID-19 Pandemic: Machine Learning Approach. J Med Internet Res 2020 Nov 25;22(11):e20550 [FREE Full text] [doi: 10.2196/20550] [Medline: 33119535]
77. Chen K, Chen A, Zhang J, Meng J, Shen C. Conspiracy and debunking narratives about COVID-19 origins on Chinese social media: How it started and who is to blame. HKS Misinfo Review 2020 Dec 10 [FREE Full text] [doi: 10.37016/mr-2020-50]

Abbreviations

AD: adjusted residual
FREX: frequency and exclusivity
Q&A: question and answer
STM: structural topic modeling

Edited by G Eysenbach; submitted 27.05.21; peer-reviewed by S Shao; comments to author 17.06.21; revised version received 20.06.21; accepted 01.08.21; published 27.08.21.

Please cite as:
Luo C, Ji K, Tang Y, Du Z
Exploring the Expression Differences Between Professionals and Laypeople Toward the COVID-19 Vaccine: Text Mining Approach
J Med Internet Res 2021;23(8):e30715
URL: https://www.jmir.org/2021/8/e30715
doi:10.2196/30715
PMID:34346885

©Chen Luo, Kaiyuan Ji, Yulong Tang, Zhiyuan Du. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Leveraging Multimedia Patient Engagement to Address Minority Cerebrovascular Health Needs: Prospective Observational Study

Elizabeth Anne Noser[1,2], MD; Jing Zhang[3,4], MPH; Mohammad Hossein Rahbar[1,3], PhD; Anjail Zarinah Sharrief[1,2], MPH, MD; Andrew David Barreto[1,2], MS, MD; Sandi Shaw[5], BSN, RN; James Charles Grotta[6], MD; Sean Isaac Savitz[1,2], MD; Nneka Lotea Ifejika[7,8,9], MPH, MD

[1]Institute for Stroke and Cerebrovascular Disease, UTHealth, Houston, TX, United States
[2]Department of Neurology, McGovern Medical School at UTHealth, Houston, TX, United States
[3]Department of Internal Medicine, Division of Clinical & Translational Sciences, McGovern Medical School at UTHealth, Houston, TX, United States
[4]Department of Biostatistics and Data Science, UTHealth School of Public Health, Houston, TX, United States
[5]Mischer Neuroscience Institute, Memorial Hermann Hospital - Texas Medical Center, Houston, TX, United States
[6]Stroke Research and Mobile Stroke Unit, Memorial Hermann Hospital - Texas Medical Center, Houston, TX, United States
[7]Department of Physical Medicine and Rehabilitation, UT Southwestern Medical Center, Dallas, TX, United States
[8]Department of Neurology, UT Southwestern Medical Center, Dallas, TX, United States
[9]Department of Population and Data Sciences, UT Southwestern Medical Center, Dallas, TX, United States

Corresponding Author:
Nneka Lotea Ifejika, MPH, MD
Department of Physical Medicine and Rehabilitation
UT Southwestern Medical Center
5323 Harry Hines Blvd
Stop 9055
Dallas, TX, 75390-9055
United States
Phone: 1 2146488762
Email: nneka.ifejika@utsouthwestern.edu

Abstract

Background: Social inequities affecting minority populations after Hurricane Katrina led to an expansion of environmental justice literature. In August 2017, Hurricane Harvey rainfall was estimated as a 3000- to 20,000-year flood event, further affecting minority populations with disproportionate stroke prevalence. The Stomp Out Stroke initiative leveraged multimedia engagement, creating a patient-centered cerebrovascular health intervention.

Objective: This study aims to address social inequities in cerebrovascular health through the identification of race- or ethnicity-specific health needs and the provision of in-person stroke prevention screening during two community events (May 2018 and May 2019).

Methods: Stomp Out Stroke recruitment took place through internet-based channels (websites and social networking). Exclusively through web registration, Stomp Out Stroke participants (aged >18 years) detailed sociodemographic characteristics, family history of stroke, and stroke survivorship. Participant health interests were compared by race or ethnicity using Kruskal-Wallis or chi-square test at an α=.05. A Bonferroni-corrected *P* value of .0083 was used for multiple comparisons.

Results: Stomp Out Stroke registrants (N=1401) were 70% (973/1390) female (median age 45 years) and largely self-identified as members of minority groups: 32.05% (449/1401) Hispanic, 25.62% (359/1401) African American, 13.63% (191/1401) Asian compared with 23.63% (331/1401) non-Hispanic White. Stroke survivors comprised 11.55% (155/1401) of our population. A total of 124 stroke caregivers participated. Approximately 36.81% (493/1339) of participants had a family history of stroke. African American participants were most likely to have Medicare or Medicaid insurance (84/341, 24.6%), whereas Hispanic participants were most likely to be uninsured (127/435, 29.2%). Hispanic participants were more likely than non-Hispanic White participants to obtain *health screenings* (282/449, 62.8% vs 175/331, 52.9%; *P*=.03). Asian (105/191, 54.9%) and African American (201/359, 55.9%) participants were more likely to request *stroke education* than non-Hispanic White (138/331, 41.6%) or Hispanic participants (193/449, 42.9%). African American participants were more likely to seek *overall health education* than non-Hispanic White participants (166/359, 46.2% vs 108/331, 32.6%; *P*=.002). Non-Hispanic White participants (48/331, 14.5%)

were less likely to speak to health care providers than African American (91/359, 25.3%) or Asian participants (54/191, 28.3%). During the 2018 and 2019 events, 2774 health screenings were completed across 12 hours, averaging four health screenings per minute. These included blood pressure (1031/2774, 37.16%), stroke risk assessment (496/2774, 17.88%), bone density (426/2774, 15.35%), carotid ultrasound (380/2774, 13.69%), BMI (182/2774, 6.56%), serum lipids (157/2774, 5.65%), and hemoglobin A_{1c} (102/2774, 3.67%). Twenty multimedia placements using the Stomp Out Stroke webpage, social media, *#stompoutstroke*, television, iQ radio, and web-based news reached approximately 849,731 people in the Houston area.

Conclusions: Using a combination of internet-based recruitment, registration, and in-person assessments, Stomp Out Stroke identified race- or ethnicity-specific health care needs and provided appropriate screenings to minority populations at increased risk of urban flooding and stroke. This protocol can be replicated in Southern US *Stroke Belt* cities with similar flood risks.

(J Med Internet Res 2021;23(8):e28748) doi:10.2196/28748

KEYWORDS

environmental justice; urban flooding; stroke; community engagement; education; health disparities

Introduction

Importance of the Problem

Social inequities affecting minority populations after Hurricane Katrina led to an expansion of the environmental justice literature after large-scale floods [1]. In 2017, Hurricane Harvey rainfall over a 4-day period was estimated as a 3000- and 20,000-year flooding event [2]. Studies after Hurricane Harvey in Houston detailed more extensive flooding among racial or ethnic minority households and those of lower socioeconomic status [3,4].

Houston is not only ranked among the most flood-impacted urban centers in the United States but also has some of the highest national stroke mortality rates, significantly affecting minority health [5]. In Harris County, Texas, which encompasses Houston, 4.7% of Medicare beneficiaries have been treated for stroke, placing Harris County in the worst quartile of all counties in the United States [6]. The combined effects of flood impact and cerebrovascular disease disparities are a growing concern for environmental justice. Correlated with the underdiagnosis and undertreatment of risk factors, minorities have a median age of first stroke 10-13 years earlier than non-Hispanic White people [5,7]. The age-adjusted incidence of first ischemic stroke is 179 per 100,000 in non-Hispanic White people, compared with 294 per 100,000 in African American people [8]. Direct medical costs for African American patients with stroke were estimated to reach 16 billion dollars in 2020; by 2030, stroke prevalence is expected to rise the most among Hispanic men, with direct costs of care increasing over 300%, compared with 140% in non-Hispanic White men, since 2012 [9].

Pros and Cons of Multimedia Stroke Prevention in Minority Populations

Minority populations are receptive to the use of mobile health (mHealth) technology for health interventions; however, racial differences in technology use and internet access persist [10]. In a recent study of stroke survivors and caregivers by Naqvi et al [11], the highest number of participants with reported internet access were non-Hispanic White people. Furthermore, studies on stroke risk factor management using mobile apps versus usual care in minority populations have yielded equivocal results [12].

Purpose

Studies have shown that ongoing public education on stroke symptoms improves stroke recognition [13,14]; unfortunately, racial disparities continue to exist regarding stroke literacy [15,16].

There are limited data regarding the practice of community engagement in stroke systems of care in flood-prone areas [17], including the method of delivery and educational activities. Furthermore, the utilization of a hybrid multimedia model—using technology to identify health care needs specific to minority populations, followed by in-person health screenings—has not been evaluated.

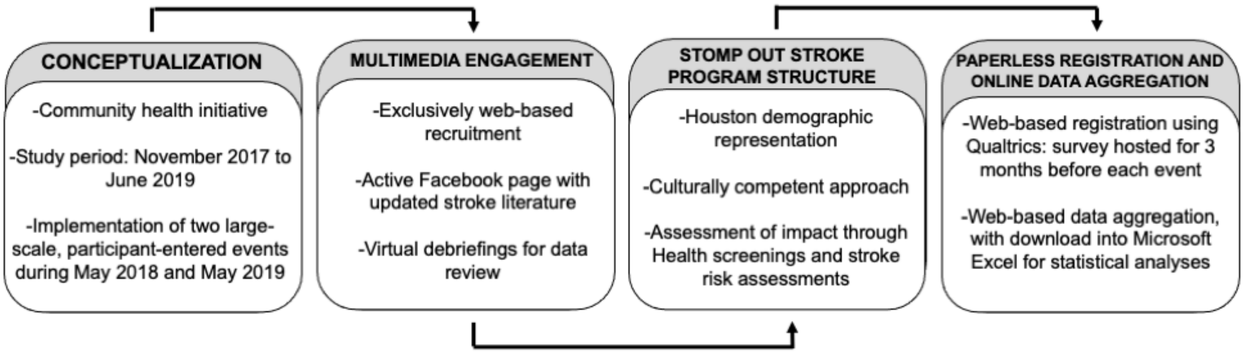
The purpose of this study was to implement Stomp Out Stroke, a hybrid multimedia education and health screening paradigm, serving a population disproportionately affected by Hurricane Harvey flooding. We hypothesize that the Stomp Out Stroke structure identifies and provides targeted health interventions, fulfilling specific needs, stratified by race and ethnicity. Multiple coastal cities in the Southern *Stroke Belt* (New Orleans, Louisiana; Charleston, South Carolina; and Savannah, Georgia) have similar demographics and flood risk, improving the generalizability of this population health intervention.

Methods

Conceptualization

Stomp Out Stroke is a prospective observational study and a collaborative public education initiative, which was implemented by the Vascular Neurology Program at the McGovern Medical School at the University of Texas Health Science Center at Houston (UTHealth) as part of the Joint Commission–Certified Integrated Stroke Healthcare System at Memorial Hermann Health System. Both institutions are located in the 2.1 square mile Texas Medical Center, topographically distributed within the 100-year and 500-year floodplains due to Brays Bayou, a slow-moving river that borders the health care district [18]. Stomp Out Stroke study period was from November 2017 to June 2019, including patient and public involvement, program structure, and implementation of two events: the first in May 2018 and the second in May 2019. Figure 1 details the Stomp Out Stroke flow process.

Figure 1. Stomp Out Stroke flow process.


CONCEPTUALIZATION
-Community health initiative
-Study period: November 2017 to June 2019
-Implementation of two large-scale, participant-entered events during May 2018 and May 2019
MULTIMEDIA ENGAGEMENT
-Exclusively web-based recruitment
-Active Facebook page with updated stroke literature
-Virtual debriefings for data review
STOMP OUT STROKE PROGRAM STRUCTURE
-Houston demographic representation
-Culturally competent approach
-Assessment of impact through Health screenings and stroke risk assessments
PAPERLESS REGISTRATION AND ONLINE DATA AGGREGATION
-Web-based registration using Qualtrics: survey hosted for 3 months before each event
-Web-based data aggregation, with download into Microsoft Excel for statistical analyses

Multimedia Engagement

Recruitment for participation in Stomp Out Stroke occurred exclusively via web-based platforms. Both the Texas Medical Center and the Texas Heart Institute provided location information, types of activities, and links to the registration website [19-21]. Stomp Out Stroke also has an active Facebook page, which is consistently updated with stroke awareness literature [22]. We used the social media hashtag *#stompoutstroke* on Twitter, Instagram, and Facebook. Members of the Vascular Neurology program also participated in television, iQ radio, and web-based news interviews to expand our reach.

Evidence-based protocols for primary stroke prevention were followed using the US Preventive Services Task Force Guide to Clinical Preventive Services [23] and the American Heart Association Guidelines for Heart-Health Screenings [24]. The Alzheimer's Foundation of America National Memory Screening program detailed the protocols for cognitive testing [25]. Virtual Stomp Out Stroke debriefings facilitated the review of data on attendee health screenings, volunteer or exhibitor evaluations, and opportunities for growth.

Stomp Out Stroke Program Structure

Houston Demographics

Houston is the fourth largest city in the United States, with a 2018 population estimate of 2.3 million people [26]. The Stomp Out Stroke program structure reflected Houston's racial or ethnic diversity and socioeconomic characteristics. The approximate racial or ethnic makeup of Houston is 45% Hispanic or Latino, 22.09% African American, 24.41% non-Hispanic White, and 6.72% Asian [27]. Overall, 30% of Houston residents were born outside of the United States. Twenty-five percent of people aged <65 years are uninsured and 20% live in poverty.

Content

Stomp Out Stroke was divided into five zones: central, stage and entertainment, healthy brain, stroke recovery, and children's. Each zone had health education and screening stations and was staffed by a volunteer coordinator and co-coordinator. Family-friendly activities and entertainment, including local multicultural dance groups, were included throughout the program (Multimedia Appendix 1).

Culturally Competent Approach

To address disparities in stroke literacy due to language barriers, bilingual health care providers and students were recruited as volunteers to assist in registration or check-in and to conduct health screenings or risk assessments in Spanish, simplified Chinese, and Vietnamese.

Assessment of Impact: Health Screenings or Stroke Risk Assessments

Assessment of impact was determined through the completion of onsite health screenings and stroke risk assessments. Each health screening consisted of vascular risk factor counseling, risk modification strategies, and recommended primary care provider follow-up. At the end of each screening, participants received printed educational materials and a give-away item (value <US $5). Automated blood pressure monitors (HEM-7222-ITZ; Omron Healthcare, Inc) were used to obtain a single measure. Additional health screenings included bone density, carotid ultrasound, BMI, serum lipids, and hemoglobin $A_{1c.}$ The 10-year stroke risk was calculated utilizing a modified Framingham stroke risk profile assessment tool, which contains age, sex, and baseline measurements of cerebrovascular risk factors, including systolic blood pressure, use of antihypertensive medications, current smoking status, presence of cardiovascular disease, current or prior atrial fibrillation, and diabetes mellitus (Multimedia Appendix 2) [28,29].

Paperless Registration and Web-Based Data Aggregation

Institutional review board approval was obtained at the McGovern Medical School at UTHealth to collect voluntary onsite and web-based registration data (HSC-MS-15-0813—UT Stroke Team Community Outreach Program). The UTHealth Institute for Stroke and Cerebrovascular Diseases has an archive of past Stomp Out Stroke events, detailing the evolution of the program before the study period [30]. During the 3 months before the event, the *Register Here* button on the front page linked participants to a web-based registration form using Qualtrics software, Version March 2020 [31]. No paper registration forms were used. Registration forms (aged >18 years) included sociodemographics, stroke survivorship, caregiver of a stroke survivor, or family history of stroke. Participants' interest in attending Stomp Out Stroke was assessed. On the event day, two check-in stations were available for preregistrants, and two stations were available for onsite registration through Qualtrics. At the end of both the 2018 and 2019 events, Qualtrics surveys were closed, and participant responses were downloaded into Microsoft Excel spreadsheets for analyses.

Statistical Analysis

Sociodemographic characteristics were summarized using frequency and percentage or median and IQR for nonnormal distributions. Comparisons by racial group or ethnicity (Asian, Black or African American, Hispanic, or non-Hispanic White) were conducted using the Kruskal-Wallis (for age) or chi-square test. A *P* value of <.05 indicated a statistically significant difference between at least two of the four ethnic groups. A Bonferroni-corrected *P* value of .0083 (.05 divided by 6) was used to control for multiple comparisons. If the pairwise comparison *P* value was <.0083, there was a statistically significant difference between multiple ethnic groups. All analyses were performed using SAS (version 9.4; SAS Institute Inc).

Results

Overview

Table 1 details the sociodemographic characteristics of Stomp Out Stroke registrants. Data from 1401 Stomp Out Stroke registrants in 2018 and 2019 were analyzed. Less than 5% of the data were missing, and imputation was not used.

Table 1. Summary of sociodemographic characteristics: 2018 and 2019 Stomp Out Stroke registrants (N=1401).

Variable	Value
Age in years (n=1381), median (IQR)	45.0 (36-57)
Sex (n=1390), n (%)	
Female	973 (70)
Male	417 (30)
Marital status (n=1383), n (%)	
Married	770 (55.68)
Not married	613 (44.32)
Race or ethnicity (n=1401), n (%)	
Hispanic	449 (32.05)
Black or African American	359 (25.62)
Non-Hispanic White	331 (23.63)
Asian	191 (13.63)
Other, two, or more races	49 (3.5)
American Indian, Alaska Native, Native American, or Pacific Islander	5 (0.36)
Unknown	17 (1.21)
Hispanic ethnicity (n=449), n (%)	
White	234 (52.12)
Black or African American	9 (2)
American Indian, Alaska Native, Native American, or Pacific Islander	8 (1.78)
Asian	3 (0.67)
Other, two, or more races	136 (30.3)
Unknown	59 (13.29)
Have children (n=1389), n (%)	
No	882 (63.5)
Yes	507 (36.5)
Health insurance (n=1342), n (%)	
Employer based	614 (45.75)
Medicare or Medicaid	251 (18.7)
Private insurance	205 (15.28)
Self-insured	44 (3.28)
Uninsured	228 (16.99)
Education (n=1331), n (%)	
More than high school	1053 (79.11)
High school or less	278 (20.89)
Are you a stroke survivor? (n=1342), n (%)	
No	1187 (88.45)
Yes	155 (11.55)
Are you a caregiver for a stroke survivor? (n=1339), n (%)	
No	1215 (90.74)
Yes	124 (9.26)
Do you have a family member who is a stroke survivor? (n=1339), n (%)	
No	846 (63.18)

Variable	Value
Yes	493 (36.81)

The median registrant age was 45 years; 69.54% (973/1390) were female and 55.68% (770/1383) were married. Overall, 32.05% (449/1401) of registrants self-identified as Hispanic, 25.62% (359/1401) self-identified as Black or African American, 23.63% (331/1401) self-identified as a non-Hispanic White person, and 13.63% (191/1401) self-identified as Asian.

A total of 63.5% (882/1389) of Stomp Out Stroke participants did not have children, 79.11% (1053/1331) had more than a high school education, and 45.75% (614/1342) had employer-based health insurance. Moreover, 155 Stomp Out Stroke participants were self-reported stroke survivors and 124 were caregivers. Nearly 36.81% (493/1339) of the registrants had a family history of stroke (Table 1).

Table 2 compares the sociodemographic characteristics of Stomp Out Stroke participants by racial or ethnic group. Hispanic participants were significantly younger than African American or non-Hispanic White participants (aged 43 years vs 48 or 49 years). A larger proportion of African American women participants were noted in comparison with Asian or non-Hispanic White people (274/358, 76.5% vs 124/188, 66% or 216/330, 65.5%).

Hispanic participants were the least likely among the four ethnic groups to have education past high school (274/432, 63.4% vs 269/322, 83.5% non-Hispanic White; 291/338, 86.1% African American; and 164/176, 93.2% Asian participants), and African American participants were the least likely to be married (128/357, 35.9% vs 200/329, 60.8% non-Hispanic White; 274/443, 61.9% Hispanic; and 123/189, 65.1% Asian participants; Table 2).

African American participants were the most likely to have Medicare or Medicaid insurance 24.6% (84/341), whereas Hispanic participants were the most likely to be uninsured 29.2% (127/435). In total, 19.9% (68/341) of African American registrants were stroke survivors; this group was the most likely to have a family history of stroke 46.5% (158/340) and more likely than non-Hispanic White or Hispanic people to have previously participated in Stomp Out Stroke (Table 2).

Table 2. Sociodemographic characteristics of Stomp Out Stroke participants by racial or ethnic group (N=1401).

Variable	Asian	Black or African American	Non-Hispanic White	Hispanic	*P* value[a]
Age (years)					<.001
Participants, n	187	353[b]	330[c]	443[b,c]	
Value, median (IQR)	45.0 (34.0-55.0)	48.0 (38.0-58.0)	49.0 (36.0-60.0)	43.0 (34.0-55.0)	
Sex, n (%)					.01
Participants	188 (100)[b]	358 (100)[b,c]	330 (100)[c]	446 (100)	
Female	124 (66)	274 (76.5)	216 (65.5)	311 (69.7)	
Male	64 (34)	84 (23.5)	114 (34.5)	135 (30.3)	
Education, n (%)					<.001
Participants	176 (100)[b,c]	338 (100)[d]	322 (100)[b,e]	432 (100)[c,d,e]	
More than high school	164 (93.2)	291 (86.1)	269 (83.5)	274 (63.4)	
High school or less	12 (6.8)	47 (13.9)	53 (16.5)	158 (36.6)	
Marital status, n (%)					<.001
Participants	189 (100)[b]	357 (100)[b,c,d]	329 (100)[c]	443 (100)[d]	
Married	123 (65.1)	128 (35.9)	200 (60.8)	274 (61.9)	
Not married	66 (34.9)	229 (64.1)	129 (39.2)	169 (38.1)	
Health insurance, n (%)					<.001
Participants	180 (100)[b,c]	341 (100)[b,d,e]	322 (100)[d,f]	435 (100)[c,e,f]	
Employer based	107 (59.4)	151 (44.3)	182 (56.5)	144 (33.1)	
Medicare or Medicaid	16 (8.9)	84 (24.6)	53 (16.5)	87 (20)	
Private insurance	22 (12.2)	50 (14.7)	49 (15.2)	70 (16.1)	
Self-insured	11 (6.1)	11 (3.2)	13 (4)	7 (1.6)	
Uninsured	24 (13.3)	45 (13.2)	25 (7.8)	127 (29.2)	
Have children, n (%)					.05
Participants	191 (100)	355 (100)	329 (100)	448 (100)	
No	135 (70.7)	225 (63.4)	220 (66.9)	270 (60.3)	
Yes	56 (29.3)	130 (36.6)	109 (33.1)	178 (39.7)	
Are you a stroke survivor? n (%)					<.001
Participants	180 (100)[b,c]	341 (100)[b,d,e]	321 (100)[c,d]	438 (100)[e]	
No	173 (96.1)	273 (80.1)	284 (88.5)	404 (92.2)	
Yes	7 (3.9)	68 (19.9)	37 (11.5)	34 (7.8)	
Are you a caregiver for a stroke survivor? n (%)					.03
Participants	178 (100)[b]	339 (100)	323 (100)[b]	435 (100)	
No	152 (85.4)	304 (89.7)	301 (93.2)	399 (91.7)	
Yes	26 (14.6)	35 (10.3)	22 (6.8)	36 (8.3)	
Do you have a family member who is a stroke survivor? n (%)					<.001
Participants	178 (100)[b]	340 (100)[b,c,d]	322 (100)[c]	435 (100)[d]	
No	117 (65.7)	182 (53.5)	211 (65.5)	292 (67.1)	
Yes	61 (34.3)	158 (46.5)	111 (34.5)	143 (32.9)	
Previously attended stroke festival? n (%)					<.001
Participants	181 (100)	343 (100)[b,c]	323 (100)[b]	439 (100)[c]	

Variable	Asian	Black or African American	Non-Hispanic White	Hispanic	*P* value[a]
No	136 (75.1)	221 (64.4)	256 (79.3)	353 (80.4)	
Yes	45 (24.9)	122 (35.6)	67 (20.7)	86 (19.6)	

[a] Except for the *P* value in the Age row, which is based on the Kruskal-Wallis test, all remaining *P* values are based on the chi-square test.

[b-f] Indicate significant differences in pairwise comparisons of either median scores or proportions in different ethnic groups. The presence of the same letter in the columns for two race or ethnic groups (ie, “a” in the Asian and “a” in the African-American race columns) indicate a significant difference between the median scores or proportions between those two respective ethnic groups.

Table 3 compares interest in attending Stomp Out Stroke by racial or ethnic group. African American registrants were more likely to attend Stomp Out Stroke for family fun and entertainment than Asian registrants (173/359, 48.2% vs 66/191, 34.6%). Hispanic registrants were more likely than non-Hispanic White people to attend Stomp Out Stroke for free health screenings (282/449, 62.8% vs 175/331, 52.9%). Asian and African American registrants were more likely to attend Stomp Out Stroke for stroke education than non-Hispanic White or Hispanic people (105/191, 54.9% Asian or 201/359, 55.9% African American vs 138/331, 41.7% non-Hispanic White or 193/449, 42.9% Hispanic). African American participants were more likely to express interest in learning about health care topics other than stroke than non-Hispanic White people (166/359, 46.2% vs 108/331, 32.6%). Non-Hispanic White people were less likely than African American or Asian registrants to speak to a health care provider (48/331, 14.5% non-Hispanic White vs 91/359, 25.3% African American or 54/191, 28.3% Asian).

Table 3. Participant interest in attending Stomp Out Stroke by racial or ethnic group (n=1330).

Variable	Asian	Black or African American	Non-Hispanic White	Hispanic	*P* value[a]
Family fun and entertainment, n (%)					.01
Participants	191 (100)[b]	359 (100)[b]	331 (100)	449 (100)	
No	125 (65.4)	186 (51.8)	188 (56.8)	270 (60.1)	
Yes	66 (34.6)	173 (48.2)	143 (43.2)	179 (39.9)	
Free health screenings, n (%)					.03
Participants	191 (100)	359 (100)	331 (100)[b]	449 (100)[b]	
No	82 (42.9)	137 (38.2)	156 (47.1)	167 (37.2)	
Yes	109 (57.1)	222 (61.8)	175 (52.9)	282 (62.8)	
Learn about stroke, n (%)					<.001
Participants	191 (100)[b,c]	359 (100)[d,e]	331 (100)[b,d]	449 (100)[c,e]	
No	86 (45)	158 (44)	193 (58.3)	256 (57)	
Yes	105 (55)	201 (56)	138 (41.7)	193 (43)	
Learn about other health care topics, n (%)					.002
Participants	191 (100)	359 (100)[b]	331 (100)[b]	449 (100)	
No	107 (56)	193 (53.8)	223 (67.4)	278 (61.9)	
Yes	84 (44)	166 (46.2)	108 (32.6)	171 (38.1)	
Speak to a health care provider, n (%)					<.001
Participants	191 (100)[b]	359 (100)[c]	331 (100)[b,c]	449 (100)	
No	137 (71.7)	268 (74.7)	283 (85.5)	353 (78.6)	
Yes	54 (28.3)	91 (25.3)	48 (14.5)	96 (21.4)	

[a] Except for the *P* value in the Age row, which is based on the Kruskal-Wallis test, all remaining *P* values are based on the chi-square test.

[b-e] Indicate significant differences in pairwise comparisons of either median scores or proportions in different ethnic groups. The presence of the same letter in the columns for two race or ethnic groups (ie, “a” in the Asian and “a” in the African-American race columns) indicate a significant difference between the median scores or proportions between those two respective ethnic groups.

During the 2018 and 2019 events, 2774 health screenings were completed within a period of 12 hours, averaging four health screenings per minute. These included blood pressure (1031/2774, 37.16%), stroke risk assessment (496/2774, 17.88%), bone density (426/2774, 15.35%), carotid ultrasound (380/2774, 13.69%), BMI (182/2774, 6.56%), serum lipids (157/2774, 5.65%), and hemoglobin A_{1c} (102/2774, 3.67%).

Multimedia Engagement

Before the 2018 event, 27 Facebook posts were displayed between December 21, 2017, and April 23, 2018, reaching 17,975 people, with 782 likes, comments, and shares; 648 post clicks; 339 page likes; and 340 page followers. The hashtag *#stompoutstroke* achieved 51 posts and 951 likes, comments, and shares on Facebook; seven posts and 30 likes, comments, and shares on Twitter; and 36 posts and 2093 likes, comments, and shares on Instagram. The Stomp Out Stroke webpage [21] achieved 18,639 page views and 9316 unique visitors. A total of 20 media placements between March 25, 2018, and April 2, 2018, reached approximately 849,731 people in the Houston area. This included 12 television stories, five web-based news stories, two iQ radio interviews, and one print story (Multimedia Appendix 3).

Discussion

Principal Findings

This study provides novel insights regarding the implementation of Stomp Out Stroke, using multimedia engagement, followed by in-person stroke education and health screening initiative, among a large minority population disproportionately affected by large-scale flooding events and cerebrovascular disease. Stomp Out Stroke registrants were representative of Houston racial or ethnic demographics, including Hispanic or Latino (449/1401, 32.05%), African American (359/1401, 25.62%), and Asian (191/1401, 13.63%). Our population was young (median age 45 years), largely female, and had received more than a high school education. Overall, 16.99% (228/1342) of Stomp Out Stroke participants were uninsured, and 18.70% (251/1342) of the participants were insured through Medicare or Medicaid.

The low adoption rates of electronic consultations for cerebrovascular risk factors, such as hypertension and diabetes, leave questions about overall use and generalizability [32]. To our knowledge, a substantial number of completed health screenings has not been replicated in the literature; indeed, the frequency of four health screenings per minute speaks to the intervention fidelity and reproducibility of Stomp Out Stroke. The multimedia impact of Stomp Out Stroke is a critical component of our hybrid model; the use of social media, hashtags, our webpage, and strategic television, iQ radio, and web-based news stories reached nearly 850,000 people.

The American Heart Association has recently focused on the use of internet-based recovery strategies for stroke survivors [33]; however, there is a paucity of research on the impact of the health and well-being of the caregiver, who may share cerebrovascular risk factors. Overall, 20.79% (279/1342) of Stomp Out Stroke registrants were stroke survivors or caregivers, a population that is undersupported by the health care system and faces unique barriers to obtaining the needed services [34]. Disabled populations are disproportionately exposed to environmental health hazards and feel abandoned by the health care system due to the marginalization of services and a lack of knowledge or skills to reengage [35]. A 2019 study of stroke survivors and caregivers enrolled in an mHealth intervention showed that female caregivers were more likely to have unknown or poorly controlled cerebrovascular risk factors [36]. The combination of multimedia engagement, followed by in-person health screening, is a hybrid model that can be used in similar urban settings.

An analysis by the City of Houston estimated that 208,353 of 848,340 households were affected by Hurricane Harvey, with a disproportionate number consisting of racial or ethnic minorities and those of lower socioeconomic status [37]. At the request of the Federal Emergency Management Agency, the National Academies of Sciences, Engineering, and Medicine appointed a committee to hold workshops in Houston post Hurricane Harvey to gain an initial understanding of the impacts of urban flooding [38]. Residents most vulnerable to flooding were again described as minorities and the poor. Houston flood district funding comes from state and federal programs, focusing on capital improvement, structural projects, home buyouts, operations, maintenance, and repairs [39]. Unfortunately, the social aspects of urban flooding are far less studied than the physical aspects; indeed, rebuilding a bridge is tangible, and building health capacity is multifaceted. Data on the intangible impact of flooding on vulnerable populations, such as stroke screenings, provide another layer of urban flood risk assessments. Future data collection and analyses could help residents of flood-prone areas receive support from civic organizations, increasing their social agency and capacity.

Community engagement paradigms focused on stroke literacy help improve the awareness of signs and symptoms [40]. We found that Hispanic populations were more likely to attend free health screenings and that African American participants were more likely to express interest in learning about nonstroke health care topics compared with non-Hispanic White people. We also identified an increased interest in speaking to a health care provider among African American or Asian participants compared with non-Hispanic White participants. Differences in the provision of health care (ie, speaking to health care providers) do not necessarily reflect deficiencies. Our goal was to provide health assessments in multiple formats. By fulfilling specific health promotion and disease prevention needs, future community engagement interventions can be tailored to the needs of culturally diverse populations.

Limitations

This study has some limitations. First, the organization and planning of Stomp Out Stroke is led by the Director of Stroke Community Outreach and Education, a full-time faculty member at the McGovern Medical School at UTHealth trained in Vascular Neurology. Her position is supported by endowment funds; sustaining Stomp Out Stroke requires collaborative efforts from multiple faculty members and neurology departmental support staff. Second, Stomp Out Stroke is funded through sponsorships, educational grants, philanthropy, and in-kind donations. Cost mitigation occurred by leveraging free educational resources at the local, state, and national levels. Third, a formalized emergency medical services and safety plan (Multimedia Appendix 4) was required for security, lost and found children, lost items, and medical emergencies. This plan was used in 2018 for a participant with transient ischemic attack symptoms. The proportion of female participants in Stomp Out

Stroke (973/1390, 70%) was higher than the female population of Houston (1,162,454/2,320,268, 50.1%), which may increase type I error when addressing the health care needs of male participants. We used the largely accepted convention that ethnicity is self-identified, based on factors such as language and shared culture. Therefore, we did not analyze the differences between the Hispanic and non-Hispanic counterparts of each race. Statistical analyses of registrants detailed the purpose of attendance and health-related needs, with no linkage to cerebrovascular risk factors or health screening data. Finally, registrants were not asked about the extent of individual losses due to Hurricane Harvey. Future studies will include zip code mapping to identify households in flood-impacted areas.

Conclusions

Stomp Out Stroke combined multimedia engagement with in-person health screenings to improve environmental justice for underserved populations at increased risk of urban flooding and cerebrovascular disease. The next step will focus on the use of mHealth technology to assess behavioral changes among repeat attendees, recurrent stroke among stroke survivors, and objective measures of stroke knowledge and preparedness.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Stomp Out Stroke program content.
[DOCX File , 651 KB - jmir_v23i8e28748_app1.docx]

Multimedia Appendix 2
Modified Framingham stroke risk profile—males and females.
[PDF File (Adobe PDF File), 180 KB - jmir_v23i8e28748_app2.pdf]

Multimedia Appendix 3
Stomp Out Stroke multimedia report.
[PDF File (Adobe PDF File), 1153 KB - jmir_v23i8e28748_app3.pdf]

Multimedia Appendix 4
Stomp Out Stroke emergency medical services and safety plan.
[DOCX File , 18 KB - jmir_v23i8e28748_app4.docx]

References

1. Colten C. Environmental justice in a landscape of tragedy. Technol Soc 2007 Apr;29(2):173-179 [FREE Full text] [doi: 10.1016/j.techsoc.2007.01.006]
2. Lindner J, Fitzgerald S. Immediate report – Final hurricane harvey - Storm and flood information - Memorandum to HCFCD Flood Watch/Partners. Flood Control District, Houston. 2018. URL: http://reduceflooding.com/wp-content/uploads/2018/06/Final-HCFCD-Harvey-Report.pdf [accessed 2017-07-16]
3. Hines E, Reid CE. Hurricane harvey hospital flood impacts: accuracy of federal emergency management agency flood hazard areas in Harris County, Texas. Am J Public Health 2020 Apr;110(4):574-579. [doi: 10.2105/AJPH.2019.305520] [Medline: 32078350]
4. Chakraborty J, Collins TW, Grineski SE. Exploring the environmental justice implications of hurricane harvey flooding in Greater Houston, Texas. Am J Public Health 2019 Feb;109(2):244-250. [doi: 10.2105/AJPH.2018.304846] [Medline: 30571302]
5. Virani SS, Alonso A, Benjamin EJ, Bittencourt MS, Callaway CW, Carson AP, American Heart Association Council on EpidemiologyPrevention Statistics CommitteeStroke Statistics Subcommittee. Heart Disease and Stroke Statistics-2020 update: a report from the American Heart Association. Circulation 2020 Mar 03;141(9):139-596. [doi: 10.1161/CIR.0000000000000757] [Medline: 31992061]
6. Stroke: Medicare population. Houston State of Health. URL: http://www.houstonstateofhealth.com/indicators/index/view?indicatorId=2067&localeId=2675 [accessed 2021-05-07]
7. Christian JB, Lapane KL, Toppa RS. Racial disparities in receipt of secondary stroke prevention agents among US nursing home residents. Stroke 2003 Nov;34(11):2693-2697. [doi: 10.1161/01.STR.0000096993.90248.27] [Medline: 14551402]
8. Kleindorfer DO, Khoury J, Moomaw CJ, Alwell K, Woo D, Flaherty ML, et al. Stroke incidence is decreasing in whites but not in blacks: a population-based estimate of temporal trends in stroke incidence from the Greater Cincinnati/Northern Kentucky Stroke Study. Stroke 2010 Jul;41(7):1326-1331 [FREE Full text] [doi: 10.1161/STROKEAHA.109.575043] [Medline: 20489177]

9. Ovbiagele B, Goldstein LB, Higashida RT, Howard VJ, Johnston SC, Khavjou OA, American Heart Association Advocacy Coordinating CommitteeStroke Council. Forecasting the future of stroke in the United States: a policy statement from the American Heart Association and American Stroke Association. Stroke 2013 Aug;44(8):2361-2375 [FREE Full text] [doi: 10.1161/STR.0b013e31829734f2] [Medline: 23697546]
10. Internet/Broadband fact sheet. Pew Research Center. 2021. URL: https://www.pewresearch.org/internet/fact-sheet/internet-broadband/ [accessed 2021-05-07]
11. Naqvi IA, Montiel TC, Bittar Y, Hunter N, Okpala M, Johnson C, et al. Internet access and usage among stroke survivors and their informal caregivers: cross-sectional study. JMIR Form Res 2021 Mar 08;5(3):e25123 [FREE Full text] [doi: 10.2196/25123] [Medline: 33683206]
12. Ifejika NL, Bhadane M, Cai CC, Noser EA, Grotta JC, Savitz SI. Use of a smartphone-based mobile app for weight management in obese minority stroke survivors: pilot randomized controlled trial with open blinded end point. JMIR Mhealth Uhealth 2020 Apr 22;8(4):e17816 [FREE Full text] [doi: 10.2196/17816] [Medline: 32319963]
13. Rimmer JH, Braunschweig C, Silverman K, Riley B, Creviston T, Nicola T. Effects of a short-term health promotion intervention for a predominantly African-American group of stroke survivors. Am J Prev Med 2000 May;18(4):332-338. [doi: 10.1016/s0749-3797(00)00129-x] [Medline: 10788737]
14. Ovbiagele B, Saver JL, Fredieu A, Suzuki S, McNair N, Dandekar A, et al. PROTECT: a coordinated stroke treatment program to prevent recurrent thromboembolic events. Neurology 2004 Oct 12;63(7):1217-1222. [doi: 10.1212/01.wnl.0000140493.83607.f1] [Medline: 15477541]
15. Ojike N, Ravenell J, Seixas A, Masters-Israilov A, Rogers A, Jean-Louis G, et al. Racial disparity in stroke awareness in the US: an analysis of the 2014 National Health Interview Survey. J Neurol Neurophysiol 2016 Apr;7(2):365 [FREE Full text] [doi: 10.4172/2155-9562.1000365] [Medline: 27478680]
16. Williams O, Teresi J, Eimicke JP, Abel-Bey A, Hassankhani M, Valdez L, et al. Effect of stroke education pamphlets vs a 12-minute culturally tailored stroke film on stroke preparedness among black and Hispanic churchgoers: a cluster randomized clinical trial. JAMA Neurol 2019 Jul 01;76(10):1211-1218 [FREE Full text] [doi: 10.1001/jamaneurol.2019.1741] [Medline: 31260028]
17. Adeoye O, Nyström KV, Yavagal DR, Luciano J, Nogueira RG, Zorowitz RD, et al. Recommendations for the establishment of stroke systems of care: a 2019 update. Stroke 2019 Jul;50(7):187-210. [doi: 10.1161/STR.0000000000000173] [Medline: 31104615]
18. Flood education mapping tool. Harris County Flood Control District. URL: https://www.harriscountyfemt.org/ [accessed 2020-04-09]
19. Stomp Out Stroke Festival 2019. Texas Heart Institute. URL: https://www.texasheart.org/event/stomp-out-stroke-festival/ [accessed 2021-05-07]
20. Stomp Out Stroke Festival offers free screenings and fun Saturday, May 4. Texas Medical Center. URL: https://www.tmc.edu/news/2019/04/stomp-out-stroke-festival-offers-free-screenings-and-fun-saturday-may-4/ [accessed 2021-05-07]
21. UTHealth Stomp out Stroke. University of Texas Health Science Center at Houston and Institute for Stroke and Cerebrovascular Disease. URL: https://www.stompoutstroke.org/ [accessed 2021-07-27]
22. UTHealth Stomp out Stroke. Facebook. URL: https://www.facebook.com/UTHealthStompOutStroke/ [accessed 2021-05-07]
23. US Preventive Services Task Force Guides to Clinical Preventive Services. Rockville (MD): Agency for Healthcare Research and Quality (US); 1996.
24. Guidelines for heart health screenings. American Heart Association. URL: https://www.heart.org/en/health-topics/consumer-healthcare/what-is-cardiovascular-disease/heart-health-screenings [accessed 2020-03-24]
25. National Memory Screening Program. Alzheimer's Foundation of America. URL: https://alzfdn.org/about-afas-national-memory-screening-program/ [accessed 2020-03-24]
26. QuickFacts, Houston city, Texas. United States Census Bureau. URL: https://www.census.gov/quickfacts/fact/table/houstoncitytexas/HSD410218#HSD410218 [accessed 2020-04-09]
27. Population and housing estimates tables. United States Census Bureau. URL: https://www.census.gov/programs-surveys/popest/data/tables.html [accessed 2020-04-10]
28. Dufouil C, Beiser A, McLure LA, Wolf PA, Tzourio C, Howard VJ, et al. Revised Framingham stroke risk profile to reflect temporal trends. Circulation 2017 Mar 21;135(12):1145-1159 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.115.021275] [Medline: 28159800]
29. Sharrief AZ, Johnson B, Urrutia VC. Stroke outreach in an inner city market: a platform for identifying African American males for stroke prevention interventions. Front Neurol 2015;6:133 [FREE Full text] [doi: 10.3389/fneur.2015.00133] [Medline: 26124744]
30. UTHealth's Stomp Out Stroke Festival - Houston's best brain day!. The University of Texas Health Science Center at Houston (UTHealth). URL: https://www.stompoutstroke.org/program-2013---2019.html [accessed 2021-05-07]
31. Qualtrics software, version March 2020. Qualtrics, Provo, UT, USA. URL: http://www.qualtrics.com [accessed 2021-07-16]
32. 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]

33. Bakas T, McCarthy M, Miller ET. Update on the state of the evidence for stroke family caregiver and dyad interventions. Stroke 2017 May;48(5):122-125 [FREE Full text] [doi: 10.1161/STROKEAHA.117.016052] [Medline: 28351961]
34. Pindus DM, Mullis R, Lim L, Wellwood I, Rundell AV, Abd Aziz NA, et al. Stroke survivors' and informal caregivers' experiences of primary care and community healthcare services - a systematic review and meta-ethnography. PLoS One 2018;13(2):e0192533 [FREE Full text] [doi: 10.1371/journal.pone.0192533] [Medline: 29466383]
35. Martinsen R, Kirkevold M, Sveen U. Young and midlife stroke survivors' experiences with the health services and long-term follow-up needs. J Neurosci Nurs 2015 Feb;47(1):27-35 [FREE Full text] [doi: 10.1097/JNN.0000000000000107] [Medline: 25565592]
36. Ifejika NL, Bhadane M, Cai C, Noser EA, Savitz SI. Cluster enrollment: a screening tool for stroke risk factors in minority women caregivers. J Natl Med Assoc 2019 Jun;111(3):281-284 [FREE Full text] [doi: 10.1016/j.jnma.2018.10.013] [Medline: 30442422]
37. Local housing needs assessment: Hurricane Harvey housing recovery. City of Houston Housing and Community Development. 2018. URL: https://houstontx.gov/housing/plans-reports/Local-Housing-Needs-Assessment-112818.pdf [accessed 2021-07-16]
38. Framing the challenge of urban flooding in the United States. In: National Academies of Sciences, Engineering, and Medicine. Washington, DC: The National Academies Press; 2019.
39. How is Harris County Flood Control District funded for disaster recovery and resiliency? Harris County Flood Control District. 2019 Oct. URL: https://www.hcfcd.org/Portals/62/About/HCFCD_FundingSources.pdf?ver=2019-12-02-151152-977 [accessed 2020-04-08]
40. Patel A, Fang J, Gillespie C, Odom E, King SC, Luncheon CA, et al. Awareness of stroke signs and symptoms and calling 9-1-1 among US adults: National Health Interview Survey, 2009 and 2014. Prev Chronic Dis 2019 Jun 20;16:E78 [FREE Full text] [doi: 10.5888/pcd16.180564] [Medline: 31228234]

Abbreviations

mHealth: mobile health
UTHealth: University of Texas Health Science Center at Houston

Edited by R Kukafka; submitted 12.03.21; peer-reviewed by L Jordan, J Gao, D Zhu; comments to author 30.04.21; revised version received 12.05.21; accepted 15.06.21; published 13.08.21.

Please cite as:
Noser EA, Zhang J, Rahbar MH, Sharrief AZ, Barreto AD, Shaw S, Grotta JC, Savitz SI, Ifejika NL
Leveraging Multimedia Patient Engagement to Address Minority Cerebrovascular Health Needs: Prospective Observational Study
J Med Internet Res 2021;23(8):e28748
URL: https://www.jmir.org/2021/8/e28748
doi:10.2196/28748
PMID:34397385

©Elizabeth Anne Noser, Jing Zhang, Mohammad Hossein Rahbar, Anjail Zarinah Sharrief, Andrew David Barreto, Sandi Shaw, James Charles Grotta, Sean Isaac Savitz, Nneka Lotea Ifejika. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Patients With Cancer Searching for Cancer- or Health-Specific Web-Based Information: Performance Test Analysis

Lukas Lange-Drenth[1], MSc; Holger Schulz[1], DPhil; Gero Endsin[2], MD; Christiane Bleich[1], DPhil

[1]Department of Medical Psychology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
[2]VAMED Rehabilitation Clinic Lehmrade, Lehmrade, Germany

Corresponding Author:
Lukas Lange-Drenth, MSc
Department of Medical Psychology
University Medical Center Hamburg-Eppendorf
Martinistraße 52
Hamburg, 20246
Germany
Phone: 49 040741056811
Email: lu.lange@uke.de

Abstract

Background: Searching the internet for cancer-related information helps patients with cancer satisfy their unmet information needs and empowers them to play a more active role in the management of their disease. However, to benefit from the search, patients need a sufficient level of skill to search, select, appraise, and apply web-based health information.

Objective: We aim to study the operational, navigational, information, and evaluation skills and problems of patients with cancer performing cancer-related search tasks using the internet.

Methods: A total of 21 patients with cancer were recruited during their stay at the rehabilitation clinic for oncological rehabilitation. Participants performed eight cancer-related search tasks using the internet. The participants were asked to think aloud while performing the tasks, and the screen activities were recorded. The types and frequencies of performance problems were identified and coded into categories following an inductive coding process. In addition, the performance and strategic characteristics of task execution were summarized descriptively.

Results: All participants experienced problems or difficulties in executing the tasks, and a substantial percentage of tasks (57/142, 40.1%) could not be completed successfully. The participants' performance problems were coded into four categories, namely operating the computer and web browser, navigating and orientating, using search strategies, and evaluating the relevance and reliability of web-based information. The most frequent problems occurred in the third and fourth categories. A total of 90% (19/21) of participants used nontask-related search terms or nonspecific search terms. A total of 95% (20/21) of participants did not control for the source or topicality of the information found. In addition, none of the participants verified the information on 1 website with that on another website for each task.

Conclusions: A substantial group of patients with cancer did not have the necessary skills to benefit from cancer-related internet searches. Future interventions are needed to support patients in the development of sufficient internet-searching skills, focusing particularly on information and evaluation skills.

(J Med Internet Res 2021;23(8):e23367) doi:10.2196/23367

KEYWORDS

telemedicine; eHealth; eHealth literacy; digital literacy; internet; web-based; health information; health education; cancer; mobile phone

Introduction

Background

Searching the internet for cancer-related information enables patients with cancer to satisfy their unmet information needs and empowers them to play a larger role in the management of their disease. Unmet information needs are one of the most frequently reported unmet supportive care needs of patients with cancer (6%-93%) during the treatment and posttreatment phases [1]. Patients with cancer prefer to receive as much information as possible about their disease [2,3]. The most common topics of cancer-related information sought on the web are information regarding the diagnosis, prognosis, disease stage, treatment options, or side effects of treatment [4-6].

The percentage of patients with cancer who use the internet to search for cancer-related information is high. In 1 Swedish, 1 American, and 1 Dutch sample, 63%-75% of the participants used the internet to search for cancer-related information or general health information [4,7,8]. The prevalence will continue to rise in the future owing to the increasing use of the internet worldwide [9].

Patients with cancer have various reasons for searching cancer-related web-based information. They use the internet to develop questions to discuss with their physician, verify information given by their physician, or seek alternative treatments [4]. Moreover, they feel that the amount of information they receive from their physician is insufficient [10].

Searching the internet for cancer-related information is positively associated with patient-reported outcomes and socioeconomic characteristics of patients with cancer. Patients with cancer who search the internet for cancer-related information are more involved in medical decision-making [11], feel better informed about their disease [7], have a higher level of self-reported health [12] and quality of life [13], are more likely to have a partner [8], and are younger and more educated [4,8,13] than patients who do not search the internet. In addition, internet health information seeking can improve the patient-physician relationship of patients with acute or chronic conditions, depending on whether the patients discuss the information with their physicians [14].

Nevertheless, to benefit from cancer-related internet searching, cancer-related web-based information must be reliable, and patients with cancer need a sufficient level of skills to search, select, appraise, and apply web-based health information [15,16]. However, the quality of cancer-related web-based information varies widely [17-23]. Information on websites is often incomplete and does not provide a basis for well-informed medical shared decision-making [17-23]. Only half (52%) of the patients with cancer trust the internet as a source of cancer-related information [24]. A total of 3 previous studies [15,25,26] analyzed essential skills to properly search the internet for health-related information. These 3 studies used performance tests and observed participants while executing health-related search tasks on the internet. The essential skills observed during task execution can be divided into 2 categories. First, people need operational and navigational skills to use a computer and web browser, that is, using a keyboard, mouse, or touch screen; navigating forward and backward between websites; and maintaining orientation on a website [15,25]. Second, they need information and evaluation skills to search, find, and assess web-based information, that is, formulate adequate search terms, choose a relevant search result, or check the source of information [15,25]. The results of the first study indicated that approximately one-third of the participants had severe problems in using operational and navigational skills [25]. Similar to the first study, the sample in the second study had, on average, a sufficient level of these skills [15]. The third study did not evaluate these 2 skills [26]. The levels of information and evaluation skills observed in the 3 studies seemed to be much lower [15,25,26]. Many participants reported problems choosing relevant search terms (14/15, 93%), selecting a reliable search result (13/15, 87%), and not checking the source of information (14/15, 93%) in at least 1 task [25]. Furthermore, none of the participants controlled the source of information, the topicality of the information, or how the information had been compiled [26].

Thus far, research on internet searching skills has focused on general healthy populations [15,26] or patients with rheumatoid arthritis [25]. The internet-searching skills of patients with cancer have not yet been studied.

Objective

Therefore, the primary goal of this study is to gain insight into the operational, navigational, information, and evaluation skills and problems of patients with cancer performing cancer-related search tasks using the internet.

Methods

Study Design

A performance test was conducted to obtain in-depth insight into the operational, navigational, information, and evaluation skills and problems of patients with cancer using the internet to search for cancer-related information on the web. Three qualitative methods of data collection were used: (1) the think-aloud method [27] combined with (2) the study administrator’s *real-time* notes and (3) video and audio data of the participants’ screen activity.

The report of this study followed the recommendations of the Standards for Reporting Qualitative Research, consisting of 21 items that aimed to improve the transparency of all aspects of qualitative research by providing clear standards for reporting qualitative research (Multimedia Appendix 1) [28].

The study protocol for this qualitative study is freely available at the Open Science Framework [29] and was published before the recruitment of the first participant.

Setting, Participants, and Recruitment

The participants were recruited during the first week of their 3-week stay at a rehabilitation clinic for oncological inpatient rehabilitation. Recruitment was conducted by the medical director (GE) of the clinic who approached the participants during the patient consultations. Patients were included if they had been diagnosed with any type of cancer and if they had sufficient oral and written proficiency in the German language. An appointment for the performance test was scheduled within the following week, and the participants received informed consent forms. Informed consent included information about the study goal, potential risks and benefits of the study, the voluntary nature of participation, and the type and duration of data storage. The participants were instructed to sign the informed consent before data collection. All appointments occurred at the rehabilitation clinic and were conducted by the same researcher (LLD). The sample size in this study was based on the concept of theoretical saturation [30], which is defined as the point when no new information, themes, or topics emerge from the data. Saturation, in the context of this study, indicates that no new performance problems were observed among the participants.

Procedure and Materials

Each appointment started with a short questionnaire to collect the following data: (1) the participants' socioeconomic characteristics (age, gender, education, and marital status); (2) their medical characteristics (cancer type, time since cancer diagnosis, and self-perceived health status measured using the second-to-last items of the German version of the EORTC QLQ C-30 [European Organization for Research and Treatment of Cancer Quality of Life Questionnaire-30] [31], with response options ranging from 1=very poor to 7=excellent); and (3) their general and cancer-related internet usage characteristics (internet experience, daily time spent on the internet, internet use for cancer-related topics, frequency of health-related internet use, and self-perceived internet searching skills). The participants had to evaluate their internet searching skills on four 5-point scales (ranging from *very bad* to *very good*) that measured the participants' self-perceived ergonomic skills, navigational skills, evaluating information reliability skills, and determining relevance skills [16].

Performance tests were started when all the items were completed. The patients executed 8 cancer-related internet search tasks (Textbox 1) based on the most common topics of cancer-related information sought on the web [4,5]. The order of the tasks was randomized for each participant because a learning effect was expected to affect the performance of subsequent tasks. The tasks were pilot tested with 4 patients with cancer to ensure comprehensibility and applicability. The participants of the pilot study were recruited from the Outpatient Clinic for Psycho-Oncology of the University Medical Center Hamburg-Eppendorf. The pilot study contained 10 search tasks. Two tasks were deleted because none of the participants were able to complete these tasks independently.

Textbox 1. Description of the cancer-related internet search tasks.

Description of the Cancer-Related Internet Search Tasks

- Imagine that you have noticed the following effects on your physical and mental well-being during your cancer treatment: listlessness, physical and mental exhaustion that does not improve even by sleep or rest. Search the internet for the symptom's name.
- Please search the internet for treatments or methods of treatment for chronic or persistent "fatigue" (this task always came after task 1).
- Please search the internet for various providers who offer psycho-oncological counseling in the Hamburg area (postcode: 22529).
- Formulate a disease-related question you have had in the past and show how you would approach this on the internet.
- Please search the patient guidelines of the German Cancer Society for your specific type of cancer.
- Search for the information sheet of the Cancer Information Service "Cancer on the internet: Surf safely."
- With the help of information from the internet, please name 5 possible side effects or symptoms of the specific cancer therapy (eg, chemotherapy and radiotherapy) that you received.
- With the help of information from the internet, please name possible ways that you could change your diet to promote your well-being or reduce side effects.

The performance tests were recorded using *Open Broadcaster Software* (version 26.1.0), which generated video and audio data. The participants were asked and trained to think aloud while performing the tasks. The verbalization of the participants' thoughts allowed the researcher to gain insight into the participants' cognitive processes while searching for web-based information [32]. In addition, the researcher present observed the participants and recorded real-time notes to identify problems with the hardware operation.

Each performance test was conducted using the same hardware (laptop, mouse, and keyboard) with identical settings. The laptop was connected to an active internet connection and was programmed with the 3 most popular web browsers (Internet Explorer, Mozilla Firefox, and Google Chrome). The participants were instructed to choose the web browser with which they had the most experience. All web browsers began with a blank page. To prevent the participants from being influenced by previous participants' search activities, the web browser was reset after each participant by removing the web browser history and cookies using CCleaner version 5.44 (Avast Software). If the participants were unable to perform the task, help was offered by the researcher present. The participants received €15 (US $17.70) for participating in the study.

Data Analysis

Statistical analyses were conducted using SPSS Statistics (version 25, IBM SPSS Inc). The participants' sociodemographic characteristics, medical characteristics, and general and cancer-related internet usage were summarized descriptively.

Video and audio data, as well as the researcher's real-time notes, were analyzed to (1) identify participants' performance problems, (2) evaluate the participants' performance, and (3) identify performance and strategic characteristics of task execution.

To identify participants' performance problems, the researchers followed an inductive coding process [33]. Participants' behavior or statements were initially coded and subsequently grouped into categories and subcategories that were then named. The category names were partly based on categories from previous research [25]. The number of problems encountered per task was then determined.

The evaluation of the participants' performance per task and the difficulty of the tasks were based on two variables: (1) could the participants complete their task *completely independently*, *with help*, or *not able to complete* the task at all, and (2) the time needed to perform the task (the more time needed to complete a task, the higher the difficulty of the task).

The execution and strategic characteristics of task execution were described by six variables: (1) the used web browser (Internet Explorer, Mozilla Firefox, or Google Chrome), (2) the starting point (eg, a specific website or a search engine), (3) the position of the opened website in the Google search listings from top to bottom (the position of the opened website in the search listings is an indicator of whether participants look beyond the first search results), (4) the number of words per search query (the use of a single search term was considered too unspecific; it is more important for the successful completion of a task to use task-related search terms than a large number of search terms), (5) the number of times a search query needed to be adjusted (a higher number of adjustments per task indicated a higher difficulty of the task), and (6) the name of the opened websites (Do the most often opened websites have a good content ranking and a commercial interest?) [21].

The influence of the participants' education (>10 years of school education vs ≤10 years of school education), age (above vs below median), self-perceived internet skills (above vs below median), internet experience (above vs below median), and time since cancer diagnosis (above vs below median) on the participants' average number of problems per task and percentage of successfully completed tasks were analyzed using 2-tailed *t* tests for independent groups. For additional interpretation, effect sizes were calculated: the values of Cohen *d* for small, medium, and large effects were 0.2, 0.4, and 0.8, respectively [34]. The α level of significance was set at α=.05.

Ethics Statement

The study was conducted in accordance with the Code of Ethics of the Declaration of Helsinki and was surveyed by the Ethics Committee of the Medical Association (Hamburg, Germany). Written informed consent was obtained from all the participants before participation.

Results

Participants' Characteristics and Participants' Internet Use

Slightly more women (12/22, 55%) than men participated in the study (Table 1). The participants' ages ranged between 25 and 81 years (mean 57 years, SD 11.9 years). Almost three-fourths of the participants lived with a partner (16/22, 73%). Most (13/22, 59%) of the sample had 10 years or less of schooling, whereas 27% (6/22) had a university degree. Breast (6/22, 27%), colon (4/22, 18%), and prostate cancer (3/22, 13%) were the most frequently reported cancer diagnoses. The participants received their diagnosis, on average, 28 (SD 57.8) months prior. The average self-perceived health status score was 4.5.

Table 1. Medical and sociodemographic characteristics of the participants (N=22).

Participant characteristics	Values
Age (years), mean (SD; range)	56.8 (12; 25-81)
Gender (female), n (%)	12 (55)
Marital status, n (%)	
Living alone	6 (27)
Living with a partner	16 (73)
Highest educational achievement, n (%)	
University degree	6 (27)
13 years of school education	3 (14)
10 years of school education	9 (41)
9 years of school education	4 (18)
Cancer type, n (%)[a]	
Breast cancer	6 (27)
Colon cancer	4 (18)
Prostate cancer	3 (13)
Lung cancer	2 (9)
Kidney cancer	2 (9)
Other	8 (32)
Time since cancer diagnosis (months), median (range)	6 (1-207)
Self-perceived health status, mean (SD)	4.5 (1.0)

[a] Multiple selection.

The participants' mean internet experience was 15 years (Table 2). Most of the participants (13/22, 59%) used the internet for less than 1 hour per day. The most common cancer-related activities on the internet were searching for cancer-related information (14/22, 64%) and communication with relatives (14/22, 64%). More than half (14/22, 64%) of the participants used the internet less than once a month for health care reasons. The participants rated their ergonomic skills, evaluating information reliability skills, navigation skills, and determining information relevance skills as *medium* to *good*.

Table 2. General and cancer-related internet usage of participants (N=22).

Participant characteristics	Values
Internet experience (years), mean (SD; range)	15.4 (7.7; 0-30)
Daily time spent on the internet (minutes), n (%)	
No utilization	1 (5)
0-30	5 (23)
30-60	7 (32)
60-120	6 (27)
>120	3 (15)
Types of internet use for cancer-related topics, n (%)[a]	
Obtaining general information about my cancer (ie, treatment information)	14 (64)
Communication with relatives or friends	14 (64)
Search for treatment options	12 (55)
Search for health care professionals	8 (36)
Verifying information received from health care professionals	7 (32)
Contact health care professionals (ie, oncologist)	4 (18)
Contact pharmacist	3 (14)
Contact other patients	3 (14)
Search for alternative treatment options	3 (14)
Search scientific data (ie, Google Scholar)	3 (14)
Frequency of internet use as a part of health care, n (%)	
Never	2 (9)
Rarely	12 (54)
More than once a month	3 (14)
More than once a week	4 (18)
Daily	1 (5)
Self-perceived internet-searching skills	
Value, range	1-5
Different self-perceived internet-searching skills, mean (SD)	
Ergonomic skills	3.5 (1.4)
Navigation skills	3.2 (1.3)
Evaluating information reliability skills	3.5 (0.7)
Determining information relevance skills	3.5 (0.8)

[a] Multiple selection.

Execution of Cancer-Related Tasks and Problems Encountered

Search Strategy and Effectiveness of Searches

Performance tests of the 21 participants were included in the analysis. The performance of participant 22 was excluded because the participant could not execute the tasks due to stress. On average, the participants executed 6.8 tasks. Participants 1 and 21 performed only 2 tasks. A total of 57% (12/21) of participants executed all 8 tasks. All data on task execution and performance problems are available in Figshare [35].

None of the participants used medical websites as a starting point. All search tasks were started using the Google search engine. On average, the participants successfully completed 59.9% (85/142) of all the tasks. Task F (Search the information sheet of the Cancer Information Service) had the highest rate of successful completions (14/18, 78%) and took participants, on average, the shortest time to execute (Table 3). Task E (Search the patient guideline for your cancer type) had the lowest rate of completion (6/19, 32%). The longest mean time (mean 323 seconds) to execute a task was observed for task D (retrieve previously searched disease information). A total of 86% (18/21%) of participants used the same web browser for all the tasks. Google Chrome was used in 80.2% (114/142) of search queries. The participants opened 192 webpages (the same website was counted every time it was opened) from 61 different websites during their 142 search queries (Multimedia Appendix 2). The 2 most frequently opened websites were provided by professional associations with generally good content rankings [21]. The participants usually selected one of the first websites from Google search listings. Of the 192 opened webpages, 163 (84.9%) webpages were ranked among the first 5 Google search results. Websites from the second page of Google search listings were opened 4 times (4/192, 2.1%). On average, the participants used 4.2 search terms per task. A single search term was used for 4.9% (7/142) of search queries. In addition, in 27.5% (39/142) of search queries, the participants decided to adjust the search terms to improve the Google search results.

Table 3. Performance and strategic characteristics of the task execution (n=21).

Tasks	A[a] (n=19)	B[b] (n=17)	C[c] (n=17)	D[d] (n=17)	E[e] (n=19)	F[f] (n=18)	G[g] (n=17)	H[h] (n=18)
Task completion, n (%)								
Completed independently	12 (63)	10 (59)	10 (59)	11 (65)	6 (32)	14 (78)	12 (71)	10 (55)
Completed with help	1 (5)	2 (12)	0 (0)	0 (0)	0 (0)	1 (6)	0 (0)	1 (6)
Not completed	6 (32)	5 (29)	7 (41)	6 (35)	13 (68)	3 (16)	5 (29)	7 (39)
Time per task (seconds)								
Value, mean (SD; range)	195 (121; 58-461)	220 (105; 68-467)	253 (207; 81-843)	323 (188; 116-696)	209 (117; 63-411)	177 (177; 44-746)	228 (146; 36-529)	235 (149; 91-489)
Web browser used, n (%)								
Google Chrome	16 (84)	15 (87)	13 (76)	13 (76)	16 (84)	14 (78)	14 (82)	13 (72)
Mozilla Firefox	3 (16)	2 (13)	4 (24)	4 (24)	3 (16)	3 (16)	3 (18)	3 (17)
Internet Explorer	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (6)	0 (0)	2 (11)
Position of website[i,j], n (%)								
First search result	11 (41)	11 (42)	6 (27)	9 (36)	15 (63)	12 (57)	7 (33)	5 (19)
Second to fifth search result	13 (48)	14 (54)	9 (41)	8 (32)	9 (37)	9 (43)	10 (48)	15 (58)
Sixth to tenth search result	3 (11)	1 (4)	7 (32)	8 (32)	0 (0)	0 (0)	4 (19)	2 (8)
Second page of Google listings	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	4 (15)
Words per search query								
Value, mean (SD)	5.6 (4.7)	3.3 (2.2)	4.8 (3.5)	3.4 (2.3)	2.8 (0.6)	3.8 (1.6)	3.4 (1.2)	3.9 (2.8)
Single search term, n (%)	0 (0)	2 (8)	1 (5)	2 (8)	0 (0)	2 (10)	0 (0)	0 (0)
Adjustments of a search query, n (%)								
Amount	6 (22)	2 (8)	5 (23)	6 (24)	4 (17)	6 (29)	6 (29)	4 (15)

[a] Identify the symptom fatigue.

[b] Search treatment methods for cancer-related fatigue.

[c] Search service providers who offer psycho-oncological counseling.

[d] Retrieve previously searched disease information.

[e] Search the patient guideline for your cancer type.

[f] Search the information sheet of the Cancer Information Service.

[g] Search for five symptoms of the treatment you received.

[h] Search for options to change your diet.

[i] Position of opened websites in the Google search listings from top to bottom.

[j] In some cases, participants opened more than one website. The position in the Google search listings of each website is listed here.

The participants' performance problems were coded into four categories concerning internet searching skills (Table 4): (1) operating the computer and web browser, (2) navigating and orientating, (3) using search strategies, and (4) evaluating the relevance and reliability of web-based information. In addition, problems in understanding the task and focusing on the task were observed.

Table 4. Performance problems and number of participants experiencing those problems for each task.

Tasks	A[a] (n=19)	B[b] (n=17)	C[c] (n=17)	D[d] (n=17)	E[e] (n=19)	F[f] (n=18)	G[g] (n=17)	H[h] (n=18)	Total[i] (n=21)
Operating the computer or web browser, n (%)									
Operating the keyboard	5 (26)	4 (24)	3 (18)	3 (18)	5 (26)	4 (22)	4 (24)	4 (22)	7 (33)
Controlling the mouse	6 (32)	2 (12)	2 (12)	2 (12)	5 (26)	5 (28)	3 (18)	6 (33)	8 (38)
Using the scroll bar	0 (0)	0 (0)	0 (0)	0 (0)	2 (11)	1 (6)	0 (0)	0 (0)	3 (14)
Operating the web browser	2 (11)	1 (6)	3 (18)	3 (18)	0 (0)	1 (6)	3 (18)	3 (17)	11 (48)
Reading difficulties	0 (0)	1 (6)	0 (0)	0 (0)	0 (0)	0 (0)	1 (6)	0 (0)	2 (10)
Participants with >1 problem per task	5 (26)	0 (0)	2 (12)	3 (18)	3 (16)	3 (18)	2 (12)	4 (22)	6 (29)
Navigating and orientating, n (%)									
Keeping orientation on a website	0 (0)	1 (6)	0 (0)	0 (0)	1 (6)	1 (6)	1 (6)	2 (11)	5 (29)
Using and understanding a PDF file	1 (5)	0 (0)	0 (0)	0 (0)	3 (11)	0 (0)	0 (0)	0 (0)	3 (14)
Using dropdown lists	0 (0)	1 (6)	0 (0)	0 (0)	1 (6)	0 (0)	0 (0)	0 (0)	2 (10)
Orientation in the Google search engine	1 (5)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (5)
Participants with >1 problem per task	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (6)	1 (6)	1 (6)	2 (10)
Using search strategies, n (%)									
Too broad search query	0 (0)	2 (12)	1 (6)	2 (12)	0 (0)	2 (12)	0 (0)	0 (0)	5 (24)
Nonspecific or nontask-related search query	9 (47)	5 (29)	9 (53)	1 (6)	5 (26)	4 (22)	4 (24)	2 (11)	19 (90)
Spelling and grammatical errors in search query	2 (11)	1 (6)	5 (29)	2 (12)	5 (26)	4 (22)	2 (12)	0 (0)	11 (52)
Adjusting the search query	1 (5)	0 (0)	1 (6)	0 (0)	2 (11)	0 (0)	1 (6)	1 (6)	6 (29)
Selection of task-related search results	0 (0)	0 (0)	3 (18)	2 (12)	0 (0)	4 (22)	4 (24)	3 (16)	11 (52)
Selecting physician instead of patient guidelines	N/A[j,k]	N/A[k]	N/A[k]	N/A[k]	8 (42)	N/A[k]	N/A[k]	N/A[k]	N/A[k]
Participants with >1 problem per task	3 (16)	0 (0)	5 (29)	0 (0)	5 (26)	2 (12)	1 (6)	1 (6)	9 (43)
Evaluating relevance and reliability, n (%)									
Controlling the source of information	19 (100)	17 (100)	17 (100)	17 (100)	18 (95)	18 (100)	17 (100)	17 (94)	21 (100)
Searching in commercial websites	1 (5)	0 (0)	2 (12)	2 (12)	0 (0)	0 (0)	3 (18)	6 (33)	8 (38)
Verifying the information	17 (89)	15 (88)	16 (94)	16 (94)	N/A[l]	N/A[l]	14 (82)	17 (94)	21 (100)
Scanning a website for relevant information	2 (11)	1 (6)	0 (0)	0 (0)	0 (0)	1 (6)	0 (0)	5 (28)	6 (29)
Participants with >1 problem per task	17 (89)	15 (88)	15 (88)	16 (94)	0 (0)	1 (6)	15 (88)	18 (100)	21 (100)
Understanding of the task and keeping focus, n (%)									
Understanding the task	4 (21)	2 (12)	0 (0)	0 (0)	1 (5)	2 (12)	3 (18)	0 (0)	7 (33)
Forgetting the task	0 (0)	2 (12)	1 (6)	1 (6)	3 (16)	3 (17)	0 (0)	0 (0)	9 (43)
Keeping focus	2 (11)	1 (6)	1 (6)	3 (18)	0 (0)	1 (6)	1 (6)	0 (0)	5 (24)

[a] Identify the symptom fatigue.

[b] Search treatment methods for cancer-related fatigue.

[c] Search service providers who offer psycho-oncological counseling.

[d] Retrieve previously searched disease information.

[e] Search the patient guideline for your cancer type.

[f] Search for the information sheet of the Cancer Information Service.

[g] Search for five symptoms of the treatment you received.

[h] Search for options to change your diet.

[i] Number of participants who experienced this problem during the execution of at least one task.

[j] N/A: not applicable.

[k] The problem was task-specific (task 5).

[l] Verifying the information was not applicable for this task because participants were instructed to search for a particular website.

Operating the Computer and Web Browser

A total of 62% (13/21) of participants had at least 1 problem using computer hardware (keyboard and mouse) or problems using basic web browser functions.

In total, 33% (7/21) of participants had problems using the keyboard, especially locating keys while typing search terms or searching for the *enter* key. In addition, 38% (8/21) of participants had problems controlling the mouse, mainly the double-clicking of icons (eg, web browser). A total of 14% (3/21) of participants experienced difficulties with the scroll bar. A total of 52% (11/21) of participants had problems operating the web browser. Frequent problems included finding the web browser icon on the desktop and closing and reopening the web browser to adjust the search terms instead of using the web browser's *back* button. Furthermore, 10% (2/21) of participants had problems reading the text of a website because of the small font size.

Overall, the operational problems were not task-specific. Most of the operating problems (94/122, 77%) were experienced by the same 6 participants. After completing several tasks, the present researcher observed that the participants became increasingly frustrated with the recurring operational problems.

In addition, 2 behaviors were observed but were not coded as operational problems. A total of 81% (17/21) of participants closed the web browser after every individual task and reopened the browser for the next task. Furthermore, 71% (15/21) of participants used a single tab for all searches.

Navigating and Orienting

A total of 33% (7/21) of participants experienced at least one problem with navigation and orientation in web browsers and on websites. Problems often occurred when the websites had complex structures, such as different graphical control elements (ie, dropdown lists or anchor links).

A total of 24% (5/21) of participants had problems maintaining their orientation on a website. They lost their orientation for various reasons. Twice, the participants felt confused by the browser's starting page (eg, "This is not where I wanted to be" [Participant 8]). Two other times, the participants did not understand the function of the anchor link at the top of the website, which would have helped them jump to the relevant information in the text. A total of 14% (3/21) of participants could not distinguish an opened PDF file from a website. In total, 10% (2/21) of participants did not find relevant information on websites because they could not use the websites' dropdown lists. One participant lost orientation due to the Google option *related searches* (eg, "I did not write that" [Participant 8]).

Notably, most of the orientation and navigation problems (94/122, 77%) were experienced by the same group of 6 participants who encountered most of the operational problems.

Using Search Strategies

A total of 95% (20/21) of participants experienced at least 1 problem using search strategies. Most of the problems occurred in the first stage when the participants formulated the search terms.

A total of 24% (5/21) of participants used only single search terms that were too broad to successfully complete the tasks. In total, 90% (19/21) of participants used nontask-related search terms or nonspecific search terms. For example, Participant 5 used the search terms *patient Hamburg* to find the patient guidelines of the German Cancer Society (task 5; nontask-related). Participant 11 searched for *psychological support* instead of *psycho-oncological support* (task 3; unspecific search). In total, 52% (11/21) of participants made spelling or grammatical errors in their search queries. These participants (9/11, 81%) usually did not adapt the search terms, and the Google option *did you mean* was not used to correct the errors.

The use of nontask-related, nonspecific, or grammatically incorrect search terms made the participants adjust their search terms or select a nontask-related webpage from Google search listings. A total of 29% (6/21) of participants experienced problems adjusting the search terms. For example, Participant 9 adjusted the original search query from *side effect of cancer therapy* to *side effect of cancer*, a query that was still not task-related (task 1: identify the symptom fatigue). A total of 52% (11/21) of participants selected nontask-related websites from the search listings. Many of these participants (5/11, 45%) randomly opened one of the first search results ("I am going on a random website; usually I don't like to pick the first website" [Participant 14]; task 2). They did not look at the URL or Google snippet (short description of the website's content) to be informed about the website's content.

To complete task 5, the participants had to find the patient guidelines (PDF file) of their specific cancer type. A total of 38% (8/21) of participants selected clinical practice guidelines instead of the patient guidelines. Both types of guidelines can be found on the same website.

Evaluating Relevance and Reliability

All participants experienced at least 1 problem while evaluating their relevance and reliability. None of the participants controlled the source or topicality of the information, except for Participant 19. Furthermore, none of the participants verified the information on 1 website with that on another website for each task. Most of the participants only opened a second website when they were not satisfied with the information on the first website. A total of 19% (4/21) participants made critical comments regarding the reliability of commercial websites; for example:

> *The first search result is an advertisement. Therefore, I will not consider that webpage.* [Participant 7]

Nevertheless, 38% (8/21) of participants selected and searched the providers' websites with a commercial interest. In total, 4 of these participants even opened websites marked as ads by Google. A total of 29% (6/21) of participants did not scan the selected websites for relevant information to complete the task. They read the websites' headings and then completed the task because they were convinced that the information they were looking for could be found on the website.

Notably, 71% (15/21) of participants made comments regarding the following: (1) the reliability of certain websites; for example:

> *I got offers. Here, from yelp.com. I have problems with opening this website because for me that is dubious information.* [Participant 17]

(2) the reliability of certain types of websites; for example:

> *What I would not read are patient forums. Where some laymen write what they did...I would rely on medical tips.* [Participant 2]

or (3) the internet as a source of cancer-related information; for example:

> *The internet in general is way too superficial. I read a book to gather information about that.* [Participant 4]

Understanding the Task and Staying Focused

A total of 33% (7/21) of participants experienced problems understanding the tasks ("I don't know what I am supposed to search" [Participant 14]; task 1), usually (10/12, 83%) resulting in nontask-related search queries. A total of 43% (9/21) of participants forgot the task during execution and had to reread it. A total of 24% (5/21) of participants were distracted by other nontask-related information. For example, Participant 3 became distracted by a brochure about *sexuality and cancer* while executing task 6.

Relationship Between Patient Characteristics and Performance Parameters

Participants 1 and 21 were not included in this analysis because they executed fewer than 3 tasks. The participants who were younger (mean 2.7, SD 1.0) and had higher self-perceived internet skills (mean 2.9, SD 1.1), on average, encountered significantly fewer (t_{17}=−2.78, P=.01, Cohen d=1.30; t_{17}=−2.33, P=.03, Cohen d=1.07) performance problems per task than those who were older (mean 4.6, SD 1.8) and had lower self-perceived internet skills (mean 4.5, SD 1.8; Multimedia Appendix 3). Both the differences had a large effect size. In addition, participants with higher self-perceived internet skills (mean 75.6, SD 18.1) completed a significantly higher percentage of tasks successfully (t_{17}=2.65; P=.02; Cohen d=1.23) than those with lower self-perceived internet skills (mean 44.9, SD 30.2). In addition, differences with a large, medium, and small effect size can be found: (1) younger participants completed a higher percentage of tasks successfully (Cohen d=0.87); (2) participants with more internet experience completed a higher percentage of tasks successfully (Cohen d=0.87) and had fewer performance problems (Cohen d=0.66) than patients with less experience; (3) participants with higher education completed a higher percentage of tasks successfully (Cohen d=0.46) and had fewer performance problems (Cohen d=0.40) than participants with a lower education; and (4) participants who had more time since diagnosis completed a higher percentage of tasks successfully (Cohen d=0.30) and had fewer performance problems (Cohen d=0.40) than participants with a lower education.

Discussion

Principal Findings

This study examined the level of operational, navigational, information, and evaluation skills of a sample of patients with cancer performing 8 cancer-related search tasks using the internet. The results indicate that a substantial group of patients with cancer did not have the necessary operational, navigational, information, and evaluation skills to benefit from cancer-related internet searches. A total of 29% (6/21) of participants had major problems with the operation of the hardware, operation of the computer and web browser, and with navigation and orientation in web browsers and on websites. These participants produced three-fourths (94/122, 77%) of the operational and navigational problems. These problems caused great frustration among the participants and often resulted in tasks not being completed successfully. A total of 6 participants completed only 29% (12/42) of their tasks successfully. Although the operational and navigational skills of most participants (15/21, 71%) seemed to be sufficient for searching the internet, the information and especially the evaluation skills were much lower. Many participants struggled with formulating a task-related search query (19/21, 90%), selecting a task-related search result (11/21, 52%) of a provider without a commercial interest (8/21, 38%), and browsing the website to find the answer to the task (6/21, 29%). Strikingly, only 19% (4/21) of participants verified the information on 1 website with that on another website, and only 5% (1/21) of participant informed himself about the provider of the website. The remaining participants seemed to take no interest in the source or topicality of the information. These findings are alarming because previous research has shown that the quality of cancer-related web-based information varies widely [17-23].

Comparison With Previous Work

Our results are consistent with those of previous studies that used performance tests to analyze the internet-searching skills of healthy participants [15,26] and patients with rheumatoid arthritis [25]. Operational and navigational problems occurred in only approximately one-third of the samples [15,25]. In addition, almost all the participants of this study and three similar studies had problems with information and evaluation skills [15,25,26]. Owing to the similarity of the results, we believe that the identified internet searching problems can also be found in patients with other health conditions and in the healthy population. In addition to the National Action Plans that plan to increase the health literacy of the entire population in the future [36,37], further web-based interventions are needed to increase the internet-searching skills of patients with low skills presently [38-40].

The rate of search queries with single search terms was the main difference between the search strategies used in 2002 [26] and those used in this study. In 2002, 65% of all search queries comprised a single search term [26] compared with 4.9% (7/142) in this study. The longer average internet experience of our sample (15.4 years vs 2.5 years) could be a possible explanation for choosing to use more search terms. However, information skills did not seem to grow with years of internet experience [41].

An exploratory analysis of our data indicated that younger age, higher self-perceived internet skills, more internet experience, and higher education were associated with encountering fewer performance problems and completing a higher percentage of tasks successfully. In addition, more time since diagnosis was associated with fewer performance problems and a slightly higher percentage of successfully completed tasks. The results of our exploratory analysis should be interpreted carefully because our sample size (n=19) was too small to make assumptions about the population of patients with cancer. However, previous research on internet-searching skills using performance tests confirmed that younger age, a higher level of education, more internet experience, and higher self-perceived internet skills are associated with more successful task completion [25,41]. Younger participants have higher operational and navigational skills but particularly poor performance regarding evaluation skills [41,42]. Education and self-perceived internet skills are associated with operational, navigational, information, and evaluation skills, whereas internet experience only has a positive influence on operational and navigational skills [41]. In addition, we analyzed the influence of time since cancer diagnosis on the participants' test performance, assuming that patients who had more time looking for cancer-related information on the web may know of more reliable providers of web-based cancer information than patients who recently received their diagnosis. A possible explanation for the lack of a stronger association with participants' performance might be that having more time since diagnosis may only be associated with patients' evaluation skills but not their operational, navigational, and information skills. Future performance tests with larger sample sizes are needed to examine this question.

Limitations

This study had several limitations. First, the participants performed the tasks in an artificial research setting under experimental conditions. They may have felt more nervous than if they had been in a natural setting. For example, Participant 22 did not start the performance test because of stress. In addition, the participants may have felt time pressure to perform the tasks and may have focused less on evaluating the reliability or topicality of the websites. We tried to minimize the pressure to perform and time pressure by explicitly reminding participants to take their time and that there were no right or wrong answers. Second, some participants had to use unfamiliar hardware as they accessed the internet exclusively with their smartphones or tablets [9]. This may have increased the number of operational problems experienced. Asking the participants what type of digital device they use to access the internet would have helped to distinguish among the users. Future performance studies should also consider enabling participants to perform tasks on a smartphone or tablet. Third, to ensure that the tasks were related to the interests and needs of patients with cancer, we formulated search tasks that covered the most common topics of cancer-related information sought on the web [4,5] and pilot-tested the tasks. However, compared with a real setting, answering our cancer-related questions had no direct effect on the treatment or well-being of the participants. Therefore, the participants may have been less motivated to evaluate the website's reliability or to verify the information found on a second website. Fourth, the study sample was small and most likely nonrepresentative. No national or international studies with representative samples have been conducted yet that allow statements on the internet searching skills of entire population groups or patients with cancer [43]. However, we expect that our sample had higher skills than the population of patients with cancer for the following three reasons: (1) the participants volunteered to participate in the study; (2) the participants' average internet experience in years (mean 15.4 years) was high, and operational and navigational skills grew with years of internet experience [41]; and (3) the participants were often highly educated (28% had a university degree), with a high level of education being related to high internet searching skills [41]. Future performance test studies should concentrate on older, lesser-educated patients with cancer with little internet experience. Additional important performance problems may be identified because these patient characteristics are associated with low internet-searching skills [25,41]. Fifth, using the concept of data saturation [30] also captures the risk of missing additional important performance problems [44]. Previous studies have indicated that even if no new concepts emerge, the possibility of further uncovered concepts in the population cannot be excluded [45,46]. We cannot completely exclude the possibility that we missed certain internet searching problems. Nevertheless, by giving the participants sufficient time to express all of their thoughts and the similarity between our results and those of previous performance tests [15,25], we are convinced that we have identified most of the important problems of patients with cancer using the internet to search for cancer-related information on the web.

Conclusions

A substantial group of patients with cancer did not have the necessary skills to benefit from cancer-related internet searches. The problems included operating the hardware, navigation and orientation in web browsers and on websites, and in particular formulating a task-related search query and critically evaluating and verifying web-based content. Given the high number of participants with higher education and relatively high internet experience, the need for future interventions or programs to increase the internet-searching skills of patients with cancer may be underestimated in this study. Additional important performance problems may be identified in future studies that concentrate on older, low-educated patients with little internet experience.

Acknowledgments

This study was financed with own funds.

Authors' Contributions

LLD was responsible for conceptualization, data curation, formal analysis, investigation, methodology, project administration, resources, validation, visualization, writing—original draft and review and editing. HS was responsible for the investigation, methodology, supervision, validation, and writing—review and editing. CB was responsible for conceptualization, investigation, methodology, supervision, validation, and writing—review and editing. GE was responsible for the recruitment of the sample, project administration, and writing—review and editing. All authors reviewed the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Standards for Reporting Qualitative Research checklist.
[DOCX File , 21 KB - jmir_v23i8e23367_app1.docx]

Multimedia Appendix 2
Names of websites and the number of times they were opened.
[DOCX File , 19 KB - jmir_v23i8e23367_app2.docx]

Multimedia Appendix 3
Completed tasks and number of encountered problems per assignment related to education level, age, self-perceived internet skills, internet experience, and time since diagnosis.
[DOCX File , 17 KB - jmir_v23i8e23367_app3.docx]

References

1. Harrison JD, Young JM, Price MA, Butow PN, Solomon MJ. What are the unmet supportive care needs of people with cancer? A systematic review. Support Care Cancer 2009 Aug;17(8):1117-1128. [doi: 10.1007/s00520-009-0615-5] [Medline: 19319577]
2. Davies NJ, Kinman G, Thomas RJ, Bailey T. Information satisfaction in breast and prostate cancer patients: implications for quality of life. Psychooncology 2008 Oct;17(10):1048-1052. [doi: 10.1002/pon.1305] [Medline: 18203242]
3. Jenkins V, Fallowfield L, Saul J. Information needs of patients with cancer: results from a large study in UK cancer centres. Br J Cancer 2001 Jan 5;84(1):48-51 [FREE Full text] [doi: 10.1054/bjoc.2000.1573] [Medline: 11139312]
4. Castleton K, Fong T, Wang-Gillam A, Waqar MA, Jeffe DB, Kehlenbrink L, et al. A survey of Internet utilization among patients with cancer. Support Care Cancer 2011 Aug;19(8):1183-1190. [doi: 10.1007/s00520-010-0935-5] [Medline: 20556435]
5. Tariman JD, Doorenbos A, Schepp KG, Singhal S, Berry DL. Information needs priorities in patients diagnosed with cancer: a systematic review. J Adv Pract Oncol 2014;2014(5):115-122 [FREE Full text] [Medline: 24910808]
6. Maddock C, Lewis I, Ahmad K, Sullivan R. Online information needs of cancer patients and their organizations. Ecancermedicalscience 2011;5:235 [FREE Full text] [doi: 10.3332/ecancer.2011.235] [Medline: 22276067]
7. van de Poll-Franse LV, van Eenbergen MC. Internet use by cancer survivors: current use and future wishes. Support Care Cancer 2008 Oct;16(10):1189-1195. [doi: 10.1007/s00520-008-0419-z] [Medline: 18293014]
8. Mattsson S, Olsson EM, Johansson B, Carlsson M. Health-related internet use in people with cancer: results from a cross-sectional study in two outpatient clinics in Sweden. J Med Internet Res 2017 May 15;19(5):e163 [FREE Full text] [doi: 10.2196/jmir.6830] [Medline: 28506959]

9. Poushter J. Social Media Use Continues to Rise in Developing Countries but Plateaus Across Developed Ones. Pew Research Center. 2018. URL: https://www.pewresearch.org/global/2018/06/19/social-media-use-continues-to-rise-in-developing-countries-but-plateaus-across-developed-ones/ [accessed 2019-12-10]
10. Chen X, Siu LL. Impact of the media and the internet on oncology: survey of cancer patients and oncologists in Canada. J Clin Oncol 2001 Dec 1;19(23):4291-4297. [doi: 10.1200/JCO.2001.19.23.4291] [Medline: 11731511]
11. Lee C, Gray SW, Lewis N. Internet use leads cancer patients to be active health care consumers. Patient Educ Couns 2010 Dec;81 Suppl:S63-S69 [FREE Full text] [doi: 10.1016/j.pec.2010.09.004] [Medline: 20889279]
12. Chou WS, Liu B, Post S, Hesse B. Health-related Internet use among cancer survivors: data from the Health Information National Trends Survey, 2003-2008. J Cancer Surviv 2011 Sep;5(3):263-270. [doi: 10.1007/s11764-011-0179-5] [Medline: 21505861]
13. Shahrokni A, Mahmoudzadeh S, Lu BT. In whom do cancer survivors trust online and offline? Asian Pac J Cancer Prev 2014;15(15):6171-6176 [FREE Full text] [doi: 10.7314/apjcp.2014.15.15.6171] [Medline: 25124593]
14. 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]
15. van Deursen AJ, van Dijk JA. Internet skills performance tests: are people ready for eHealth? J Med Internet Res 2011 Apr 29;13(2):e35 [FREE Full text] [doi: 10.2196/jmir.1581] [Medline: 21531690]
16. 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]
17. Al-Bahrani A, Plusa S. The quality of patient-orientated internet information on colorectal cancer. Colorectal Dis 2004 Sep;6(5):323-326. [doi: 10.1111/j.1463-1318.2004.00604.x] [Medline: 15335363]
18. Borgmann H, Mager R, Salem J, Bründl J, Kunath F, Thomas C, et al. Robotic prostatectomy on the web: a cross-sectional qualitative assessment. Clin Genitourin Cancer 2016 Aug 1;14(4):e355-e362. [doi: 10.1016/j.clgc.2015.12.020] [Medline: 26821529]
19. Bruce JG, Tucholka JL, Steffens NM, Neuman HB. Quality of online information to support patient decision-making in breast cancer surgery. J Surg Oncol 2015 Nov;112(6):575-580 [FREE Full text] [doi: 10.1002/jso.24046] [Medline: 26417898]
20. Hargrave DR, Hargrave UA, Bouffet E. Quality of health information on the internet in pediatric neuro-oncology. Neuro Oncol 2006 Apr;8(2):175-182 [FREE Full text] [doi: 10.1215/15228517-2005-008] [Medline: 16533758]
21. Liebl P, Seilacher E, Koester M, Stellamanns J, Zell J, Hübner J. What cancer patients find in the internet: the visibility of evidence-based patient information - analysis of information on German websites. Oncol Res Treat 2015;38(5):212-218. [doi: 10.1159/000381739] [Medline: 25966768]
22. Nghiem AZ, 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]
23. Wasserman M, Baxter NN, Rosen B, Burnstein M, Halverson AL. Systematic review of internet patient information on colorectal cancer surgery. Dis Colon Rectum 2014 Jan;57(1):64-69. [doi: 10.1097/DCR.0000000000000011] [Medline: 24316947]
24. Lange L, Peikert ML, Bleich C, Schulz H. The extent to which cancer patients trust in cancer-related online information: a systematic review. PeerJ 2019;7:e7634 [FREE Full text] [doi: 10.7717/peerj.7634] [Medline: 31592341]
25. van der Vaart R, Drossaert CH, de Heus M, Taal E, van de Laar MA. Measuring actual eHealth literacy among patients with rheumatic diseases: a qualitative analysis of problems encountered using Health 1.0 and Health 2.0 applications. J Med Internet Res 2013 Feb 11;15(2):e27 [FREE Full text] [doi: 10.2196/jmir.2428] [Medline: 23399720]
26. 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]
27. Ericsson KA, Simon HA. Verbal reports as data. Psychol Rev 1980;87(3):215-251. [doi: 10.1037/0033-295X.87.3.215]
28. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014 Sep;89(9):1245-1251 [FREE Full text] [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
29. Lange L, Schulz H, Bleich C, Endsin G. How do cancer patient search for cancer or health specific online information? A qualitive analysis of navigation strategies and problems when using the internet. Open Science Framework. 2019. URL: https://osf.io/mqu8z [accessed 2019-08-01]
30. Guest G, Bunce A, Johnson L. How many interviews are enough? Field Methods 2016 Jul 21;18(1):59-82. [doi: 10.1177/1525822X05279903]
31. Waldmann A, Schubert D, Katalinic A. Normative data of the EORTC QLQ-C30 for the German population: a population-based survey. PLoS One 2013;8(9):e74149 [FREE Full text] [doi: 10.1371/journal.pone.0074149] [Medline: 24058523]

32. Hak T, van der Veer K, Ommundsen R. An application of the three‐step test‐interview (TSTI): a validation study of the dutch and norwegian versions of the ‘illegal aliens scale’. International Journal of Social Research Methodology 2006 Jul;9(3):215-227. [doi: 10.1080/13645570600656264]
33. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
34. Cohen J. Statistical Power Analysis for the Behavioral Science. Hillsdale, NJ: Routledge; 1988.
35. Lange-Drenth L. How do cancer patients search for cancer- or health-specific online information. Figshare. 2020. URL: https://figshare.com/articles/journal_contribution/How_do_cancer_patients_search_for_cancer-_or_health-specific_online_information_/12770510/1 [accessed 2020-08-06]
36. Baur C, Harris L, Squire E. The US national action plan to improve health literacy: a model for positive organizational change. Stud Health Technol Inform 2017;240:186-202. [Medline: 28972518]
37. Schaeffer D, Hurrelmann K, Bauer U, Kolpatzik K, Gille S, Vogt D. [National action plan health literacy: need, objective and content]. Gesundheitswesen 2019 Jun;81(6):465-470. [doi: 10.1055/a-0667-9414] [Medline: 30273942]
38. Robinson C, Graham J. Perceived Internet health literacy of HIV-positive people through the provision of a computer and Internet health education intervention. Health Info Libr J 2010 Dec;27(4):295-303 [FREE Full text] [doi: 10.1111/j.1471-1842.2010.00898.x] [Medline: 21050372]
39. Xie B. Effects of an eHealth literacy intervention for older adults. J Med Internet Res 2011 Nov 3;13(4):e90 [FREE Full text] [doi: 10.2196/jmir.1880] [Medline: 22052161]
40. Mitsuhashi T. Effects of two-week e-learning on eHealth literacy: a randomized controlled trial of Japanese Internet users. PeerJ 2018;6:e5251 [FREE Full text] [doi: 10.7717/peerj.5251] [Medline: 30013857]
41. van Deursen AJ, van Dijk JA, Peters O. Rethinking Internet skills: the contribution of gender, age, education, Internet experience, and hours online to medium- and content-related Internet skills. Poetics 2011 Apr;39(2):125-144. [doi: 10.1016/j.poetic.2011.02.001]
42. Gui M, Argentin G. Digital skills of internet natives: different forms of digital literacy in a random sample of northern Italian high school students. New Media Soc 2011 Mar 8;13(6):963-980. [doi: 10.1177/1461444810389751]
43. Karnoe A, Kayser L. How is eHealth literacy measured and what do the measurements tell us? A systematic review. Know Manag Learn 2018;7(4):576-600. [doi: 10.34105/j.kmel.2015.07.038]
44. 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]
45. Perneger TV, Courvoisier DS, Hudelson PM, Gayet-Ageron A. Sample size for pre-tests of questionnaires. Qual Life Res 2015 Jan;24(1):147-151. [doi: 10.1007/s11136-014-0752-2] [Medline: 25008261]
46. O'Reilly M, Parker N. 'Unsatisfactor saturation': a critical exploration of the notion of saturated sample sizes in qualitative research. Qual Res 2012 May 17;13(2):190-197. [doi: 10.1177/1468794112446106]

Abbreviations

EORTC QLQ C-30: European Organization for Research and Treatment of Cancer Quality of Life Questionnaire-30

Edited by R Kukafka; submitted 14.08.20; peer-reviewed by A Hidki,Dr, P Postigo-Martin; comments to author 04.10.20; revised version received 16.10.20; accepted 24.05.21; published 16.08.21.

Please cite as:
Lange-Drenth L, Schulz H, Endsin G, Bleich C
Patients With Cancer Searching for Cancer- or Health-Specific Web-Based Information: Performance Test Analysis
J Med Internet Res 2021;23(8):e23367
URL: https://www.jmir.org/2021/8/e23367
doi:10.2196/23367
PMID:34398801

©Lukas Lange-Drenth, Holger Schulz, Gero Endsin, Christiane Bleich. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 16.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Investigating the Use of Telemedicine for Digitally Mediated Delegation in Team-Based Primary Care: Mixed Methods Study

Michael Knop[1], MSc, RN; Marius Mueller[1], PhD; Bjoern Niehaves[1], PhD

[1] Chair of Information Systems, University of Siegen, Siegen, Germany

Corresponding Author:
Michael Knop, MSc, RN
Chair of Information Systems
University of Siegen
Kohlbettstrasse 15
Siegen, 57072
Germany
Phone: 49 15755910502
Email: michael.knop@uni-siegen.de

Abstract

Background: Owing to the shortage of medical professionals, as well as demographic and structural challenges, new care models have emerged to find innovative solutions to counter medical undersupply. Team-based primary care using medical delegation appears to be a promising approach to address these challenges; however, it demands efficient communication structures and mechanisms to reinsure patients and caregivers receive a delegated, treatment-related task. Digital health care technologies hold the potential to render these novel processes effective and demand driven.

Objective: The goal of this study is to recreate the daily work routines of general practitioners (GPs) and medical assistants (MAs) to explore promising approaches for the digital moderation of delegation processes and to deepen the understanding of subjective and perceptual factors that influence their technology assessment and use.

Methods: We conducted a combination of 19 individual and group interviews with 12 GPs and 14 MAs, seeking to identify relevant technologies for delegation purposes as well as stakeholders' perceptions of their effectiveness. Furthermore, a web-based survey was conducted asking the interviewees to order identified technologies based on their assessed applicability in multi-actor patient care. Interview data were analyzed using a three-fold inductive coding procedure. Multidimensional scaling was applied to analyze and visualize the survey data, leading to a triangulation of the results.

Results: Our results suggest that digital mediation of delegation underlies complex, reciprocal processes and biases that need to be identified and analyzed to improve the development and distribution of innovative technologies and to improve our understanding of technology use in team-based primary care. Nevertheless, medical delegation enhanced by digital technologies, such as video consultations, portable electrocardiograms, or telemedical stethoscopes, can counteract current challenges in primary care because of its unique ability to ensure both personal, patient-centered care for patients and create efficient and needs-based treatment processes.

Conclusions: Technology-mediated delegation appears to be a promising approach to implement innovative, case-sensitive, and cost-effective ways to treat patients within the paradigm of primary care. The relevance of such innovative approaches increases with the tremendous need for differentiated and effective care, such as during the ongoing COVID-19 pandemic. For the successful and sustainable adoption of innovative technologies, MAs represent essential team members. In their role as mediators between GPs and patients, MAs are potentially able to counteract patients' resistance toward using innovative technology and compensate for patients' limited access to technology and care facilities.

(J Med Internet Res 2021;23(8):e28151) doi:10.2196/28151

KEYWORDS

digital health; digital health care technologies; telemedicine; user perceptions; delegation; primary care; ambulant health care; medical assistants; general practitioners; COVID-19; mixed method study; multidimensional scaling; mobile phone

Introduction

Background

The current health care systems are facing major challenges. Several shortcomings are prevalent with regard to the availability of medical professionals and facilities, which impedes the provision of comprehensive care. In particular, rural areas are undergoing a rapid demographic change, leading to higher patient numbers and increased occurrences of age-related health issues, which results in a higher health care demand [1]. Simultaneously, the numbers of general practitioners (GPs) and specialized physicians have been decreasing in these areas [2], in part because they often find it more attractive to establish themselves in bigger cities [3]. Thus, doctors often struggle to find successors who are willing to take over their practice, which in turn leads to higher workloads for the remaining GPs who must meet the growing demand [4]. Consequently, an inequitable distribution of health care services has emerged that disadvantages rural areas that are often structurally weaker [5,6]. Nowadays, these disadvantageous trends are further complicated by the ongoing COVID-19 pandemic, which has led to governmental decisions restricting personal contacts among society as well as between patients and health care providers [7]. Therefore, a transition away from in-person treatment can be observed, evoking novel challenges that need to be addressed to maintain comprehensive access to care [8,9], such as privacy and cybersecurity concerns [10,11], the potential conduct of inaccurate examinations, or the undermining of patient-physician relationships [8].

As a reaction and countermeasure to these challenges, new care models have emerged that alter the structures and delivery processes in health care and seek to free up resources and enable GPs to cope with increasing demands and contemporary restrictions [12]. One example of a novel way of organizing comprehensive ambulant patient treatment is the deployment of medical assistants (MAs) who are entitled to an advanced set of permissions and responsibilities [13-15]. MAs are meant to take on some of the GP’s tasks, such as conducting home visits or adjusting medications (which are done in collaboration with the GP). Accordingly, these new structures and processes call for new ways of communicating, documenting, and practicing care that account for the multiple actors [16]. Although GPs need to be empowered to delegate some of their duties to the assistants to free up their own capacities and thus be able to cope with increasing demands, MAs must form ways to warrant their work, align it with medical standards and routines, and thus contribute to effective and safe treatments. Hence, new collaborative forms of care emerge, and responsibilities are partially disseminated across professionals [13]. However, this multi-actor approach also presents challenges, such as patient compliance with these novel processes. In addition, the importance of effective communication increases within team-based primary care, as MAs are not allowed to treat patients autonomously because of legal boundaries. In this regard, treatment errors and communication problems between actors can occur that call for generating digital competencies at an early stage [17] and the willingness to participate in this digital transformation.

In addition to approaches dealing with the prevalent issues in health care that involve innovative changes in personnel management and delegation, the application of digital technologies within care delivery and treatment processes has also been shown to be effective under specific conditions. Technologies such as telemedical video consultation systems for efficient patient-physician communication [18], body-worn sensory equipment that allows for patient-sided collection of vital data [19], and telemonitoring systems that render in-person contact unnecessary [20] have been applied to bridge gaps in patient treatment and in the availability of medical professionals in the workforce. In this context, the emerging internet of things brings together a variety of data collected by users and ubiquitous connected devices such as biosensors and smart meters [21,22]. Further applications of these technologies include virtual home visits [23], remote examinations [24], digital prescriptions [25], and scheduling appointments [26] and the provision of information on diseases, symptoms, and possible treatments that can be easily accessed on the web [27]. In addition, internet of things apps can be used to support older or chronically ill patients at home, thus contributing to an independent way of life [21]. In particular, with regard to the ongoing COVID-19 pandemic, telemedicine (such as video consultations and telemonitoring) has been shown to enable spatially independent treatment while ensuring quality of care and patient safety [28]. Nevertheless, research has also shown that the digitalization of primary health care processes might lead to contrary effects and requires careful consideration of underlying conditions. Research argues that digital-first approaches to general practice might lead to an increase in workload without sufficient differentiation of patients and their needs [29].

Drawing a synopsis of these two perspectives, the integration of digital technologies and new care delivery structures holds the potential to further improve health care quality and comprehensiveness, with the ultimate goal of maintaining or even improving the safety, satisfaction, and overall health of the patient. Research delivers initial insights in that regard, showing that the integration of technologies into multi-actor health care processes can yield higher allocative efficiency and organizational outcomes (eg, lower hospitalization rates) [30]. These targeted benefits are of particular interest when looking at rural areas and the prevailing circumstances, such as a lack of work force and resulting per capita demand [31]. Here, the application of digital technologies within multi-actor health care processes holds the potential to address the challenges and issues that are vital to address, such as more complex communication and delegation paths or the need for dispersed and transparent accountability [32]. Using digital technologies, GPs and MAs alike are potentially able to collect richer data on their patients based on measured vital parameters or user inputs. It can be assumed that the augmentation of innovative, multi-actor care delivery models with supportive digital technologies represents a promising and more holistic strategy, which calls for dedicated studies examining whether prevalent challenges in health care and structural disparities, such as in rural and remote areas, can be approached in a beneficial and feasible way.

Multi-Actor Approaches in Modern Health Care

The shortage of medical professionals in Western countries has led to the development of multiple strategies to counter the difficulties related to the provision of medical services in primary care [33], especially through shifting from clinical to ambulant care [34] and through medical task delegation, that is, the transfer of medical interventions from a doctor to another medical professional [33]. Although task delegation in Western countries normally involves the transfer of medical interventions from doctors to nurses [35,36], GPs also delegate tasks to MAs in several countries [37]. Although *medical assistant* is a general term for several professionals with different medical skills and training, MAs are usually part of a GP's staff [15]. The former role of MAs in physicians' offices was focused on administrative tasks and the provision of isolated clinical measures under the direct supervision of physicians [38]. Currently, MAs are qualified to visit patients at home and manage different tasks, for example, monitoring a patient's health status, taking blood samples, or supervising and adjusting the intake of prescribed medication. Although MAs are usually not permitted to diagnose patients or adjust medical treatments on their own, the role of MAs increasingly comprises the more complex task of continuously assessing and evaluating a patient's health status to ensure optimal case management together with primary care physicians [39]. Undertaking tasks including patient education, health promotion, or monitoring the social and psychological well-being of patients, the role of MAs and primary health care nurses share specific similarities and might converge in the future [40]. Nevertheless, the differentiation and definition of the roles MAs and nurses take for primary care require further investigation and are not always clear [12,41]. Through the delegation of tasks in ambulant primary care to MAs, GPs are able to better distribute and structure their workload, resulting in a more effective and satisfactory work routine [33]. Therefore, the success of multi-actor approaches in health care depends on specific structural, organizational, and outcome-associated conditions, for example, efficient communication between team members [42] or patient satisfaction [43].

Bridging Gaps in Primary Care Through Digital Technologies

Considering these challenges of multi-actor approaches in primary care, the digitalization of health care technologies improves the ability to catalyze team-based, multidisciplinary, and resource-sensitive processes. Considering the uniqueness of multi-actor approaches in primary care, the technologies that are relevant are those that are capable of enhancing collaboration, communication, documentation, and patient intervention. Information and communication technologies (ICTs) in particular have the potential to optimize multi-actor care processes [44], for example, by enabling remote access to patient information or documenting care through digital platforms [45]. For digitally mediated delivery of care, a wide range of telemedicine systems provide different services, from audiovisual applications for digital appointments (video consultation) [46,47] to sensory-enhanced systems for auscultation (listening to the internal sounds of the patient, usually with a stethoscope) [48]. In addition, body-worn sensors or other monitoring devices allow autonomous and continuous measurements to generate more accurate, rather than isolated, medical data [19,49]. For digitally mediated health care, it is of great interest to differentiate between the potential of different types of digital technologies and to explore factors that are crucial for their adoption and sustainable integration into existing work routines [50]. Therefore, the role of digital technologies in multi-actor approaches is bound to the evolving informational gaps that are caused by new structures and processes in primary care.

Objectives

To date, the literature lacks studies that shed light on the potential and benefits of combining multi-actor care processes and supportive digital technologies in primary care. Hence, this study seeks to investigate the attitudes, perceptions, expectations, and needs of medical professionals located in a rural area, who play a role in multi-actor patient treatment processes. To accomplish this, our research draws upon qualitative results gathered from semistructured interviews as well as results from a web-based survey that was completed by both GPs and MAs. Both interviews and surveys were conducted in a region characterized by rural conditions and associated challenges. This mixed methods approach allows for a triangulation of findings and delivers richer insights into the target groups' attitudes and perceptions regarding the use of digital technologies in health care, thus shedding further light on how these health care actors cope with increasing efforts provoked by the rural environment. Consequently, the objectives of this study are to investigate (1) the technologies that are suitable and effective for application in multi-actor care delivery and delegation processes in rural care and (2) the factors underlying the professionals' use and perception of identified technologies that are already in use or exhibit future applicability.

Methods

Study Design

We conducted a three-step, mixed methods approach to thoroughly investigate the phenomena of interest. As part of a regional project in Rhineland-Palatinate, Germany, with 11 different primary care physicians' offices, this study empirically explored the potential of various digital technologies for enhancing delegation processes in rural primary care. Following our initial research objectives, we (1) collected and analyzed qualitative data from 19 interviews with GPs and MAs from 11 different primary care physicians' offices participating in the project. Through our process of analysis, we discovered that the perceived differences between technologies that were already being used by medical professionals and new innovative technologies seemed to be important factors for the hypothetical adoption or rejection of innovative technologies. Therefore, to expand our understanding of the perception of innovative digital technologies by GPs and their MAs, we (2) conducted a web-based survey within the same population and used multidimensional scaling (MDS), which will be further explained later within this section, to reveal underlying patterns of technological preferences. GPs and MAs from our sample confirmed the results of our MDS in a subsequent workshop.

Finally, we achieved a richer and deeper understanding of the investigated phenomena by (3) triangulating the results of both data sets [51,52].

Interview Study

Overview

As part of a regional project on the digitalization of delegation processes in German primary care, we conducted 19 qualitative interviews with GPs and MAs in 11 different rurally situated offices (we ensured to conduct at least one interview with participants from each office). All MAs that participated in this study underwent basic nonacademic clinical training for 3 years and had a supplementary qualification enabling them to undertake clinical tasks in ambulant care, comprising 190-270 hours of training. Originally, we had planned to conduct individual interviews only. Nevertheless, some offices asked us to conduct group interviews because of the high workload and time pressure. Therefore, we conducted 13 individual interviews and 6 group interviews with comparable characteristics. Table 1 provides an overview of the interviewees' characteristics.

Table 1. Summary of interviewees' characteristics.

Characteristic	Type of interview		
	Group	Individual	Total
Interviews, n (%)	6 (32)	13 (68)	19 (100)
Duration (minutes), mean (SD)	77 (23.4)	59 (16.0)	68 (20.0)
Participants, n (%)	13 (50)	13 (50)	26 (100)
Age (years), mean (SD; range)	43 (11.2; 26-59)	49 (9.2; 31-61)	46 (10.5; 26-61)
Job experience (years), mean (SD; range)	20 (9.1; 7-35)	26 (7.3; 9-37)	23 (8.6; 7-37)
Profession, n (%)			
Medical assistant	8 (57)	6 (43)	14 (100)
General practitioner	5 (42)	7 (58)	12 (100)
Gender, n (%)			
Male	4 (57)	3 (43)	7 (100)
Female	9 (47)	10 (53)	19 (100)

In addition to demographic characteristics and general questions about their profession, we asked participants about (1) their current organizational processes of delegating medical services in ambulant care, the role of (digital) technologies for these processes, and possible solutions for emerging difficulties; (2) their relationship to patients and how (digital) technologies shape or affect them; and (3) the reciprocity of self-perception and the use of (digital) technologies, that is, how participants' own understanding of their professional role affects their attitude toward (digital) technologies. The full interview guidelines can be found in Multimedia Appendix 1. Note that not every question was asked during each interview. If a direct or indirect answer was given before the respective question was asked, it was skipped by the interviewer. In this way, we were able to reduce redundancy in the data and provide a more streamlined interviewing experience. Through these semistructured interviews, we intended to recreate the daily routines or processes of GPs and MAs regarding medical delegation and their use of technology to moderate or facilitate these routines. Subsequently, participants were asked about hypothetical scenarios involving the use of innovative digital technologies in the near or distant future, for example, video consultation or automated monitoring of medical parameters, such as blood pressure, blood coagulation, or blood glucose levels. Following previous theoretical and empirical work, we plan to outline insights into the use of innovative technologies to facilitate and assist delegation processes, as well as insights into the GPs' and MAs' subjective perspectives and understanding of technological effectiveness.

Interview Data Collection and Analysis

The regional project in Germany, which included this study, involved 11 GPs' offices in rural areas. Following a purposeful sample [53], we included all 11 offices by conducting at least one interview (group or individual) with GPs or MAs from each office. The participating staff members from each office were contacted and interviewed by 2 different members of the research group (MK and MM). As identifying the differences between the perspectives of GPs and MAs on the use of innovative digital technologies in primary care was considered an important goal of this study, we focused on individual interviews with representatives of each profession. All interviews were conducted face to face between August and October 2019. In addition, we conducted a smaller number of focus group interviews. We did not change our aforementioned guidelines to ensure comparability throughout all interviews. Interviews were audio recorded, transcribed nonverbatim, and translated into English. As we were primarily interested in content-related insights, we did not conduct a sequential analysis and left out pauses and emotional or nonverbal sounds (such as sighs or laughter) from the transcription. The interviewees signed an informed consent form before the start of their interview.

To analyze the qualitative data, we followed a three-fold approach that was applied in a previous study in the health care domain [31]. First, 2 authors from the research group (MK and MM) independently coded each interview. On the basis of the grounded theory methodology proposed by Strauss and Corbin [54], the coding process comprises open, axial, and selective coding, which is described subsequently. This approach is particularly useful in this study because it performs well when inductively analyzing qualitative (unstructured) data with little or no previous knowledge. Hence, the highly explorative approach chosen in this study, combined with the lack of pre-existing research and insights about the phenomenon under investigation, qualified the grounded theory methodology as a fitting analysis paradigm. Following this procedure, each author started with open coding by intuitively assigning in vivo codes to the interview texts. Where possible, the first categories were formed by subsuming related open codes. Then, after going through each interview, axial codes were formed by categorizing open codes into broader schemes, thus achieving a higher level of abstraction. Next, superordinate, selective codes were formed that representatively subsumed related or redundant axial codes, which represented the theoretical core findings on a top level of abstraction. Therefore, two independent coding schemes emerged, each comprising selective, axial, and open codes. Second, in a process of comparison, the 2 authors discussed their coding schemes with regard to the research objectives and dissolved disagreements in code formulation and categorization, meaning, and code-to-text assignments. This step produced a reconciled coding scheme consisting of three overarching categories (ie, selective codes) that represent the essential findings of our qualitative analysis. In the third and final step, both authors each recoded the data by applying the novel scheme, followed by a conclusive discussion and approval of the coding procedure.

Survey Study

Overview

Owing to inconsistencies in the findings from the qualitative study, which are discussed later in the *Results* section of our qualitative findings, we decided to conduct a second data collection to explore latent dimensions of subjective technology valuation by GPs and MAs. Research shows that the combination of semistructured interviews and MDS appears to be a valuable approach to gain a deeper understanding of the differences in participants' subjective perceptions or underlying beliefs [55,56]. Therefore, we reached out to participants through a web-based survey conducted in March 2020. The survey consisted of demographics and a sorting question that asked participants to bring 10 different digital technologies into a hierarchical order following their perception of how relevant these technologies are or would be for their everyday work. The named technologies were derived inductively from our qualitative study and represented technologies that are already used for medical delegation by all offices or discussed purely for future use within our sample of 11 different offices. Therefore, these 10 technologies represent a combination of innovative digital technologies for primary medical care and technologies that were already in use by GPs and MAs. Table 2 lists the included technologies along with a short definition. Please note that the status of use in Table 2 (*technologies already in use*) reflects the time before or just at the beginning of the COVID-19 pandemic. Status of use was determined by the findings of our qualitative study and is defined by the categories *Yes* (technologies are used by all offices) or *No* (technologies are not used by any office).

To visualize and analyze nonmetric sorting data to build clusters and interpret the latent dimensions of valuation, we used MDS. Through this approach, we were able to further examine the aforementioned contradictions from our qualitative findings.

Table 2. Definition and summary of technologies used in the web-based survey to explore general practitioners' and medical assistants' latent dimensions of technology use.

Type of technology	Abbreviation used for analysis	Characteristics	Technologies already in use
Telemedical electrocardiogram (12-lead)	TelECG	Records electronic signals of a patient's heart to assess the cardiac health status of a patient with the same quality as a stationary 12-point ECG[a], but allows remote operation at a patient's home through MAs[b] and real-time data transmission to a physician	No
Electronic medical record	EMR	Collects and stores patient data electronically; helps to organize and structure medical care in clinical settings	Yes
Blood pressure monitor	BPM	Measures the blood pressure of a patient to assess information about a patient's cardiac or general health status; can be either electronic or manual	Yes
Blood coagulation monitor	BCM	Measures the coagulation level of a patient's blood; often used on patients taking medication to thin their blood after cardiac or neurological incidents	Yes
Telemedical stethoscope	TelSteth	Instrument to auscultate heart, lungs, or other body parts of a patient like a stethoscope, but allows remote operation at a patient's home through MAs and real-time data transmission to a physician	No
Mobile venoscope	MobVen	A mobile instrument to detect veins and venation through transillumination; facilitates the puncture of veins	No
Smartphone	SP	Mobile phone with computer-like functions, including verbal and text-based communication, internet access, camera use, navigation, and its own operating system	Yes
Blood glucose monitor	BGM	Measures the level of glucose in patient's capillary blood; mainly used on patients with metabolic diseases, especially diabetes	Yes
Digital appointment	DA	Audiovisual appointment (video consultation) between patient and general practitioner or patient and MA to digitally assess a patient's health status	No
Infrared thermometer	InfTherm	Measures the body temperature of a patient; indicates inflammatory processes in a patient's body, for example, infections	Yes

[a] ECG: electrocardiogram.

[b] MA: medical assistant.

Survey Data Collection and Analysis

As this study aimed to determine the underlying dimensions of perception participants had to judge the value of a specific technology for their work, we contacted participants from our qualitative study and asked them to participate in our additional web-based survey. Therefore, we ensured a purposeful sampling approach within the same population of participants to draw conclusions from both data sets appropriately. Therefore, 2 members of the research group (MK and MM) contacted participants from the qualitative study who were then asked to participate in the survey study. The participants started the survey by providing their consent. From the initial 26 GPs and MAs participating in our qualitative interviews, 14 (54%) responded to our web-based survey, from which three data sets had to be removed because of incomplete data. The demographics of the remaining 11 participants are summarized in Table 3.

MDS originates from psychological research, which has been used as an explorative approach for determining latent dimensions of nonmetric or metric data. It has been applied in different contexts regarding technology use, such as to differentiate between types of e-marketplaces [57] or to address technology-related phenomena in organizational research [58,59] and, more recently, to explore latent dimensions for context-specific technology use, such as cyberdeviance [60]. Owing to its applicability to behavioral phenomena and its ability to operate with small sample sizes [61], MDS was selected as an applicable statistical method to enrich our insights from the qualitative interviews.

Table 3. Demographic information about participants from the web survey.

Characteristics	Values
Gender, n (%)	
Male	4 (36)
Female	7 (64)
Profession, n (%)	
Medical assistant	5 (45)
General practitioner	6 (55)
Age (years), mean (SD; range)	45 (12.3; 27-62)
Job experience (years), mean (SD; range)	21(9.1; 9-34)
Patient size of physician’s office (number of patients treated per year), mean (SD; range)	4940 (1876.3; 1000-7600)

In an MDS configuration, objects are represented as points in a multidimensional space (usually 2D or 3D). The distances between the points correspond to the empirical dissimilarity of the objects [62]. Therefore, MDS transfers dissimilarity data into a geometrical configuration with m dimensions ($m \in N$). Thereby, dissimilarity data can be represented by pairwise ratings of objects, intercorrelations, or hierarchically sorted preferences [62]. Through a specific mathematical algorithm, dissimilarity data are then used to generate an optimized one- or multidimensional configuration representing the starting point for theoretical reasoning. For our second data set consisting of technologies that participants sorted by their subjective perception of relevance for their daily work routines, we used a special type of MDS called multidimensional unfolding (MDU) [62,63]. MDU creates an optimized geometrical representation of participants, together with their subjective perceptions of technologies, enabling a visual interpretation of the latent dimensions of these perceptions. We used the R Foundation’s statistical software R [64] with the package smacof [65] for statistical analysis. The geometrical configuration was calculated using the Euclidean distance function to define the distances between included objects. Euclidean distances were calculated using the formula for Minkowski distances d_{ij} (X) while setting the distance’s order p=2.

Eucledean distances allow an intuitive interpretation of geometrical configurations, as they represent our natural understanding of distances [63]. For instance, the Euclidean distance between two points in a 2D space is determined by a direct line between these points. In an MDU configuration, small distances represent high similarities between objects and vice versa. The dimensional suitability of our statistical solution was evaluated by comparing the *stress−1* values for different numbers of dimensions [62,66]. *Stress−1* values yield a normed indicator for the differences between the distances within the actual configuration d_{ij} (X) and a function $f(p_{ij})$ with distances , called disparities, representing the ideal model for the empirical data [63].

Compared with error functions from other statistical methods, such as regression analysis, a minimization of the *stress−1* value (convergence toward 0) is desirable. Furthermore, we evaluated the number of iterations required to calculate the configuration.

Results

Results From Semistructured Interviews (Interview Study)

Overview

From our qualitative interviews with GPs and MAs, three main categories emerged for discussing the potentials of digital technologies in facilitating delegation processes: direct patient treatment, documentation and communication of treatment, and contrast of personal interaction and telemedicine. The main categories represent the highest level of content-related interpretation, following our approach for qualitative data analysis.

Direct Patient Treatment

Our first category reflects participants’ perspectives on the important role of digital technologies for medical diagnosis and treatment through delegation. While recreating their daily work routines, GPs and MAs primarily described technologies that they already use to take direct measurements of medical parameters (mainly blood glucose levels, blood pressure, and blood coagulation levels) of patients in ambulant medical care. When considering the ongoing digitalization of the measurement of vital parameters (eg, continuous automated monitoring of blood glucose levels or the automated data transfer of blood pressure measures to a patient’s GP), participants evaluated present and future scenarios based on accuracy and instantaneousness of measurement, as well as the usability of technologies. For the measurement of blood pressure, GPs and MAs primarily use a combination of blood pressure cuff and stethoscope, both manual technologies. As digital versions of these measurement tools are easier to use but do not provide the same accuracy, GPs and MAs continue to use manual technology:

> *Sometimes, a patient has horrible blood pressure measurements [while using a digital blood pressure monitor], so you go to visit him. Then you measure it [blood pressure of the patient] yourself and tell the patient to put a new battery in it or something like that, so maybe it will take a better measurement next time.* [Interview 16, GP]

Concerning the measurement of blood coagulation levels, both GPs and MAs mentioned the importance of technological innovation for the feasibility of delegation processes in ambulant care and time-efficient treatment processes. In the past, GPs or MAs had to take blood samples through venous punctuation and send them to a medical laboratory to determine a patient's coagulation level. Digital versions of blood coagulation monitors can now instantly and accurately measure the parameters using capillary blood. Therefore, digital coagulation monitors are currently used almost exclusively for ambulant care:

> *In the past, . . . GPs had to visit the patients and had to take blood samples for coagulation, even before consultation hours. In the afternoon they had to see the patients again . . . it was very complicated, and now it's easy, isn't it?* [Interview 15, MA]

Similar to blood coagulation levels, the blood glucose levels of patients are also usually measured with digital equipment, as only capillary blood is needed for an instant analysis and accuracy does not vary much in comparison to venous sampling. GPs mentioned that MAs were competent enough to assess whether a patient's blood pressure, blood glucose level, or blood coagulation level necessitated notifying the GP about the patient's health status. While discussing potential digital innovations for these kinds of technologies, such as automated data transfer or push notifications in case of an unusual deviation in a patient's vital parameters, GPs preferred receiving the subjective interpretation from an MA through a direct phone call, as GPs were likely to ask additional questions about a patient in a potential emergency case:

> *They [MAs] give me a call when something's wrong. . . . and they're very quick—quicker than some typed message that I maybe wouldn't hear. In this situation, automated data transfer isn't of any use. When a patient has a blood glucose level of 60 [hypoglycemia], they know what to do; they know it's too low and they have to do something.* [Interview 14, GP]

Aside from already familiar technologies, participants discussed in detail the potential of two innovative digital technologies for ambulant medical care: mobile telemedical electrocardiograms (ECGs) and telemedical stethoscopes. In contrast to monitors for blood glucose levels, blood pressure, and blood coagulation levels, the interpretation of ECGs or auscultation sounds is highly complex and is usually not delegated to MAs. However, MAs are competent in recording patient data for the GP:

> *I can't evaluate it [ECG], that's the doctor's business . . . I can put it on a patient . . . and then I bring it to the office. There it's evaluated and the doctor decides what to do with it.* [Interview 15, MA]

Therefore, a mobile version of an ECG that is able to transfer an ambulant patient's data in real time to the GP's office (hereafter called a telemedical ECG) was discussed in the interviews as an innovative technology that could be used in daily work. Mostly, GPs and MAs believed that this technology would be helpful in recording and analyzing a patient's ECG data remotely. Especially when an ambulant patient's medical issues occur spontaneously, MAs liked the idea because they could reassure themselves and the patient through a direct evaluation of the ECG data:

> *To transfer ECG data, that would be great. I'd like that a lot, I could imagine, to somewhat delegate medical problems* . . . [Interview 2, GP]

Nevertheless, accuracy remained an important factor for GPs and MAs in deciding to actually use a telemedical ECG, as they emphasized the necessity of a telemedical ECG to generate a quality of medical data that are comparable with state-of-the-art stationary ECGs (12-lead). Although participants found the idea of a telemedical ECG highly interesting and relevant, the GPs were especially pessimistic about the cost-effectiveness. Owing to the high purchase prices and the lack of reimbursement by social health insurance, numerous GPs formulated resistance to actually purchasing a telemedical ECG:

> *If I had one [telemedical ECG], that would be really helpful, but you have to consider that I couldn't even charge something for the use of it. . . . So alas, it's a cost-benefit analysis once more.* [Interview 18, GP]

Similarly, the opportunity to auscultate ambulant patients from a distance by instructing qualified MAs on how and where to put a digital stethoscope on the body of a patient seemed to be of high interest for GPs. By enabling the collection of medical data and real-time transfer to a GP's office, telemedical stethoscopes were seen by GPs as a likely way to save time:

> *Well, it surely would make my work easier, if I don't need to visit every older patient with a cold anymore, if I could just auscultate them remotely.* [Interview 19, GP]

GPs and MAs mentioned that a telemedical stethoscope could be practical because of its variety of uses. Telemedical ECGs are used solely for medical issues involving functional cardiac abnormalities. Telemedical stethoscopes' indications include cardiac, pulmonic, and unspecific medical problems.

Documentation and Communication of Treatment

Our second category contains the participants' discussions of the potentials arising from digital documentation and communication. GPs and MAs emphasized the importance of mobility and ubiquity of information, but also mentioned hindrances for the sustainable use of digital technologies. While discussing the digitalization of communication about patients and their treatment processes, GPs and MAs pointed out the role of direct contact between GP and MAs in daily work routines. As MAs might unexpectedly encounter a patient with severe medical issues, all of the included GPs' offices provided an emergency call system. Thus, MAs can talk to a GP at any time if they think it is necessary. In addition, participants talked about the use of private messaging systems for cases in which communication was not urgent. As one of the most recent widely discussed innovations in primary care, digital telemedicine systems that enable audiovisual communication (video consultation), combined with the remote transfer of medical data in real time (eg, telemedical stethoscopes) were perceived as helpful for delegation processes by GPs and MAs. Three different benefits were primarily associated with telemedicine systems: replacing a GP's home visit through digitally mediated delegation, improving (ad hoc) diagnostics in ambulant medical care, and reassuring MAs in ambulant medical care. Although GPs were skeptical about fully replacing direct bilateral contact with a patient through telemedicine, telemedicine-mediated interactions involving MAs were considered helpful in some situations:

> *If I could actually have a look at the patient, like in a video conference, and our medical assistant asks some additional questions and I got some relevant parameters, I could really imagine saying, "Well, the patient's all right; he's stabile, so there's no need for me to visit him."* [Interview 2, GP]

In addition, GPs mentioned that they would not rely on a patient's ability to use a telemedicine system for diagnostic purposes on their own; for example, as the use of telemedical ECGs might be complex, and potential misuse might lead to misleading information. Therefore, the medical competence of MAs was discussed as an important factor for using digital technologies to potentially improve ambulant medical care and overcome a patient's lack of competence in adequately using such technology. Telemedicine systems were considered by MAs to be useful as an innovative channel of communication that can help to better determine the potential diagnosis of a patient after they initially assess the patient's health status:

> *If I could send some data directly to the doctor, that would make things easier. So that telemedicine, when I put on an ECG, the doctor might tell me from his desk, if it's alright or not, if we have to call an ambulance or if he needs to visit the patient by himself.* [Interview 10, MA]

Through telemedicine systems, MAs saw the possibility of contacting a GP regarding a medical situation that they were not fully able to assess. Closely related to the improvement of the quality of a diagnosis in delegation processes, telemedicine seemed to provide a feeling of security and reliability:

> *When facing a critical situation, I think it [video consultation] could help me to feel safe. Because I'm usually alone on-site . . . I think I would be more confident when I think the patient's not looking good, I better turn it on and the doctor sees what I see.* [Interview 6, MA]

Although telemedicine systems, therefore, appear to be helpful for direct communication in ambulant medical care, health records are used to collect and sort patient information and data over a longer period of time. Participants discussed the use of mobile electronic health records (EHRs), and MAs especially pointed out the importance of having access to relevant patient health records in ambulant medical care. As direct contact with a GP is usually reserved for critical situations, MAs discussed their need for a medium through which to collect relevant information about a patient's health status. Although all participating offices installed some type of EHR in their systems, remote access to patient information seemed to be difficult, as mobile versions of EHR were not installed on suitable devices or their actual use was impractical. Nevertheless, participants emphasized the potential benefits of easy-to-use mobile documentation that could make manual documentation obsolete:

> *. . . and then I write it down, write it all down and sometimes I can't figure out my handwriting afterwards and therefore it [mobile EHR] would be useful. So you can write it down directly and be connected to a patient's entire medical history.* [Interview 18, MA]

Although some participants mentioned that manual documentation was carried out very quickly while making home visits, most participants reported the necessity for double documentation later:

> *. . . here [doctor's office], you're sitting for hours to write down everything you did on home visits. It takes a lot of time. So, if you could just make your documentation while you're still on home visits, that would save a lot of time.* [Interview 9, MA]

Furthermore, GPs remarked on the lack of interoperability, not only between mobile software apps and applications in their office but also between their own documentation software and the software used by hospitals, nursing homes, or other GPs. Especially when patients move from clinical to ambulatory care, GPs pointed out that transferring the patient's clinical data or updating medication takes a lot of time because the inability to transmit records electronically means that it must be done manually:

> *When a patient comes to me with his clinical reports, that's a catastrophe. You get hand-written reports, . . . you have to transfer into the EHR. Also, it doesn't work with ambulant nursing care, it doesn't work with nursing homes.* [Interview 8, GP]

Communication and documentation technologies seem to be essential factors for a dynamic and uninterrupted workflow, from the perspectives of GPs and MAs. Although telemedicine systems enabling audiovisual communication between the GP, MA, and patient or real-time remote transfer of medical data (ECG and auscultation sounds) are recognized as helpful digital innovations, the digital technologies that are already in use (EHR) do not seem to have reached their full potential because of a lack of interoperability and user-friendly mobile apps.

Contrast of Personal Interaction and Telemedicine

In our third category, we subsumed participants' discussion of the risks and limitations of digital technologies concerning team-based primary care and the interaction between them and patients. Aside from considerations about the usefulness of digital technologies regarding direct patient interventions or superordinate processes of medical care, participants perceived the use of technology as contextual, that is, directly or indirectly embedded in a specific social interaction. Although the positive aspects of telemedicine system use were discussed, most participants were skeptical about replacing direct physical interaction with digitally mediated interaction:

> *It makes a difference . . . It's not the same. I think the gold standard is a direct encounter, to be in the same room; that's just different to [audio-visual telemedicine].* [Interview 4, GP]

Participants further differentiated the limitations of telemedicine use into those resulting from a perceived restriction of relevant patient data (eg, skin conditions, walking behavior, or general appearance) and those resulting from the absence of bodily contact itself. In particular, GPs emphasized their need to use multiple sensory inputs to correctly diagnose a patient:

> *Well, I don't know how he [the patient] smells, I can't have a look at his skin. Is he sweating? Is his skin cold not supplied with enough blood? I want to examine him [physically]. And these are things that are most relevant for the diagnosis you find, in the end. For me, it plays an important role.* [Interview 7, GP]

Some MAs argued that not only is the ability to evaluate the physical conditions of a patient reduced by digitalized home visits but also the ability to evaluate environmental factors, such as the general condition of a patient's home or the presence of objects that could potentially increase the fall risk in older patients:

> *Also, it's about fall risk; I have an eye on that. There might be a new carpet, causing a risk or cables lying in a patient's way. I talk to the patient about these things, or his relatives.* [Interview 8, MA]

Aside from more objective limitations, GPs and MAs discussed the meaning of bodily contact with a patient as a part of the social interaction itself. Participants mentioned the importance of direct contact with patients, not just to medically treat them correctly and comprehensively but also to form a relationship with them. Participants were not able to completely explain what underlying assumptions lead them to the impression that a direct, nondigitally mediated interaction is preferable to the use of telemedicine for the purpose of social interaction. Nevertheless, they emphasized the advantage of direct contact with the patient to build a relationship in which trust can be created and patients feel safe to talk about personal problems:

> *The bodily presence. The contact . . ., especially for older patients. My job is especially about old people. They need it. Or maybe some joking or something like that. You won't do that when you're on the computer. When you're sitting directly with each other, then some things are discussed. And that's missing while using telemedicine. If everything was to be digital, something would be missing.* [Interview 10, MA]

In contrast, participants also considered the social effects that the use of technology itself has on patients, for example, the feeling of security and control when blood pressure or blood coagulation are being monitored. In particular, GPs compared the positive effect of technological use on patients with the placebo effect known from medication:

> *The patient has a good feeling, then. Technology is always great. Something beeps, some additional measurement for some specific symptoms. Technology has something like a placebo effect.* [Interview 5, GP]

In addition, participants assumed that technology adoption by patients was highly affected by the attitude of the attending GP or MA. Participants reasoned that the technological adoption of patients might depend on the formulated medical necessity and the explanation given by the GP or MA:

> *But they don't really ask many questions about it [long-term blood pressure monitor or ECG]. We tell them how it is done, how it works. And then it's all right, so it's very uncommon that someone asks questions about it. They rely on what we said about it.* [Interview 10, MA]

In summary, GPs and MAs were aware of the factors that influence the relationship between them, the patient, and the use of technology. Participants reflected on the potential effects of innovative technologies on social interaction and discussed the limitations from their point of view. Interestingly, our three main categories, therefore, represent three different dimensions of technological use in ambulant medical delegation: (1) an interventional dimension involving direct patient contact, defining the action of care; (2) a superordinate dimension of communicating and documenting care; and (3) a reflective social dimension in which participants discussed contextual and relational factors of technology use.

In the process of analyzing and summarizing the qualitative findings from the focus group and individual interviews, contradictions were identified between the interviewees' perceptions of innovative technologies and technologies that were already in use. Several times during the interviews, participants seemed to reject innovative technologies (eg, telemedicine systems used to contact patients from a distance) while using specific arguments (eg, risk of private data misuse) that they did not apply to technologies that were already in use (eg, smartphone use to communicate patient data). As the underlying mechanisms of subjective technology valuation must be considered an important factor for technology use [67], in an effort to answer our initial research questions, we decided to conduct our survey study to explore latent dimensions of subjective technology valuation by GPs and MAs.

Results From MDU (Survey Study)

By applying MDU, we explored the underlying dimensions of the GPs' and MAs' perceptions of the usefulness of relevant technologies in facilitating the process of delegation in medical ambulant care. Qualitative results suggest the various categories and factors that GPs and MAs use to evaluate different types of technologies; the subjective motivation and individual perception of technologies help to fully understand and explain the behavioral intention and the actual use of technology by medical professionals. By merging the results from both the qualitative and survey studies, we intended to contribute to a comprehensive understanding of the role of digital technologies in primary care delegation processes.

To determine the number of dimensions that are appropriate for interpretation, we analyzed the *stress−1* values as well as the number of iterations for different MDU configurations (Table 4). Although values for *stress−1* decreased with greater numbers of dimensions, iterations for dimension m=4 reached its maximum of 10,000 and were excluded after visualization from potential solutions because of its obvious triviality. For all calculated configurations, we followed the suggestions of de Leeuw et al [65] and Busing et al [68] to avoid degenerate solutions.

As it showed the lowest *stress−1* value, we selected the configuration with m=3 for further analysis and interpretation. As a *stress−1* value of 0.18 represents an approximately correct solution, we tested the reliability of the configuration in relation to the data by calculating a random solution for m=3 (number of iterations=1) and compared it with the initial configuration [63]. The random solution had a *stress−1* value of 0.38. As the *stress−1* value of our initial 3D configuration appears to be much smaller [63], we considered our result to be satisfactory. Figure 1 shows a 3D depiction of the resulting configuration.

To provide a possible interpretation of our results, we changed the angles of our selected configuration and fixed one dimension after another (Figures 2-4) using the R package scatterplot3d [69]. Demographic control variables (eg, age or gender) did not lead to a sufficient explanation of the following dimensions and were therefore not discussed as potential criteria for interpretation.

Table 4. Summary of different configuration characteristics.

Dimension	Number of iterations	*stress−1* value
1	12	0.41
2	62	0.24
3	113	0.18
4	10,000	0

Figure 1. 3D configuration of technologies and participants to visualize the perceived technological relevance for daily work routines (numbers represent participants, words represent different technologies). BCM: blood coagulation monitor; BGM: blood glucose monitor; BPM: blood pressure monitor; DA: digital appointment; EHR: electronic health records; InfTherm: infrared thermometer; SP: smartphone; TelECG: telemedical electrocardiogram; TelSteth: telemedical stethoscope.

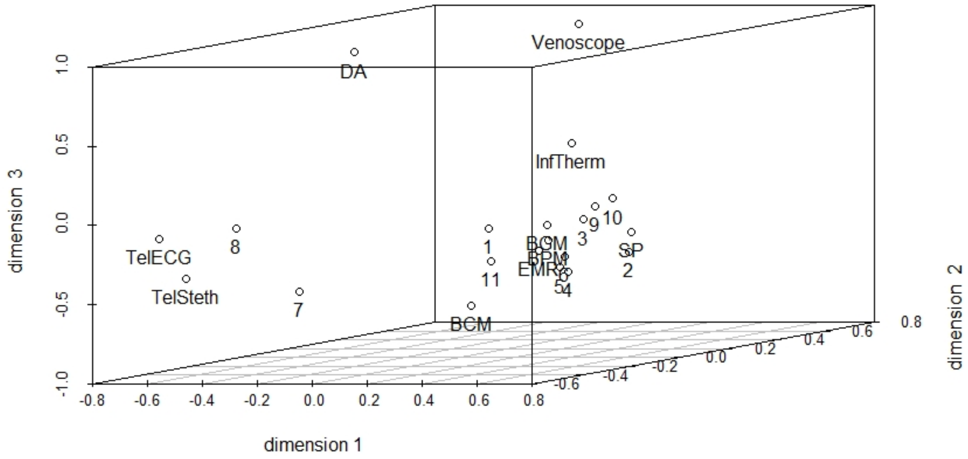


Figure 2. 3D configuration with fixed first dimension (numbers represent participants, words represent different technologies). BCM: blood coagulation monitor; BGM: blood glucose monitor; BPM: blood pressure monitor; DA: digital appointment; EHR: electronic health records; InfTherm: infrared thermometer; SP: smartphone; TelECG: telemedical electrocardiogram; TelSteth: telemedical stethoscope.

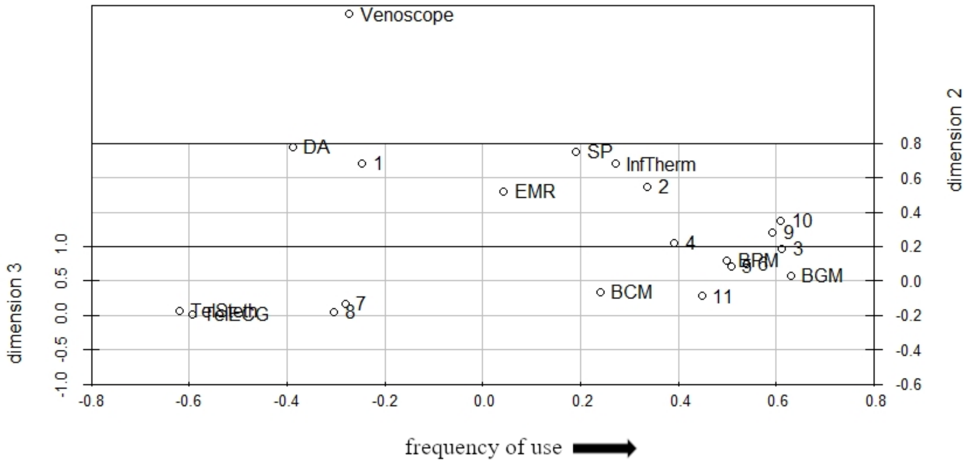


Figure 3. 3D configuration with fixed second dimension (numbers represent participants, words represent different technologies). BCM: blood coagulation monitor; BGM: blood glucose monitor; BPM: blood pressure monitor; DA: digital appointment; EHR: electronic health records; InfTherm: infrared thermometer; SP: smartphone; TelECG: telemedical electrocardiogram; TelSteth: telemedical stethoscope.

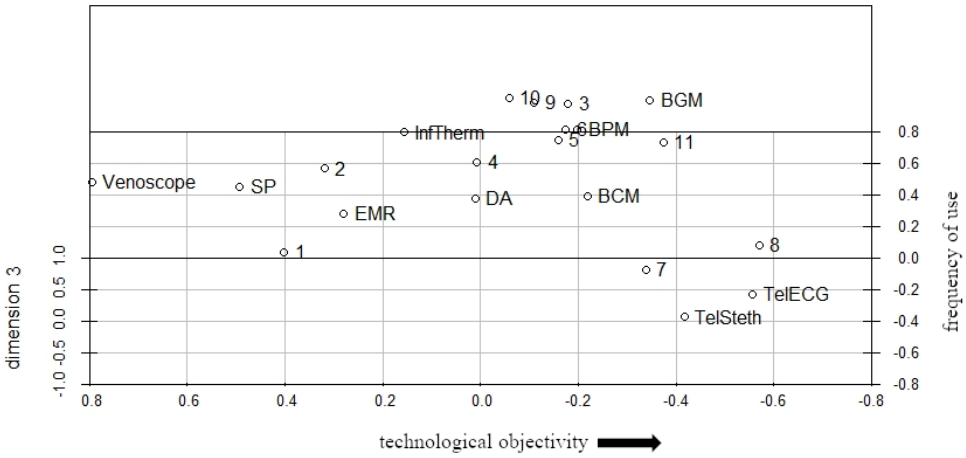
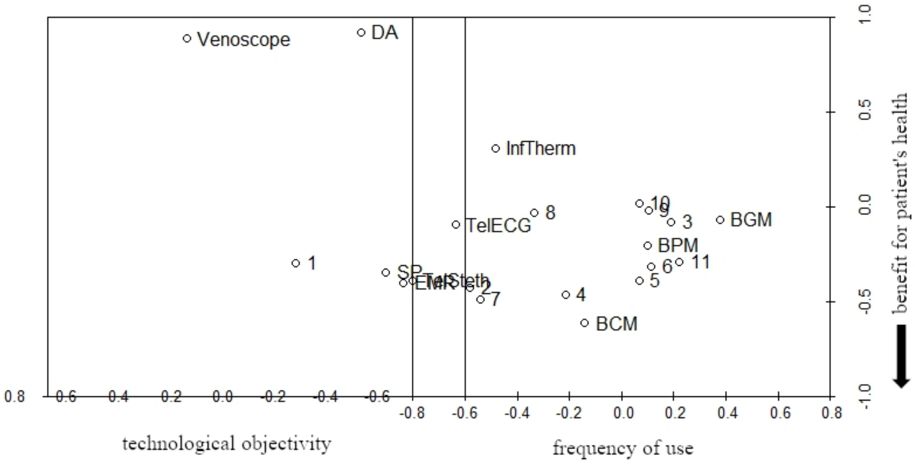


Figure 4. 3D configuration with fixed third dimension (numbers represent participants, words represent different technologies). BCM: blood coagulation monitor; BGM: blood glucose monitor; BPM: blood pressure monitor; DA: digital appointment; EHR: electronic health records; InfTherm: infrared thermometer; SP: smartphone; TelECG: telemedical electrocardiogram; TelSteth: telemedical stethoscope.



In Figure 2, we propose the underlying dimension *hypothetical* and *actual frequency of use*: innovative digital technologies that were not yet in use at the time of data collection, such as telemedical stethoscopes and ECGs or video consultation, are sorted to the left of the dimension *frequency of use*, whereas frequently used technologies, such as blood pressure or blood glucose monitors, are sorted to the right. Notably, the venoscope appears to be near innovative digital technologies that were not yet in used at the time of data collection, as GPs and MAs of our sample usually do not use it in ambulant medical care.

Therefore, the perceived relevance of technological use seems to be strongly associated with daily work routines and the possible incorporation of specific technologies. Although most of the participants are sorted to the right, meaning that they tend to find the technologies that they already use to be relevant, participants 1, 7, and 8 are sorted near innovative digital technologies. Therefore, most participants take into account their current behavioral routines when evaluating the relevance of a specific technology, whereas a minority of participants forecast a frequent use of innovative technologies whose characteristics and benefits they only assume but do not have experience with.

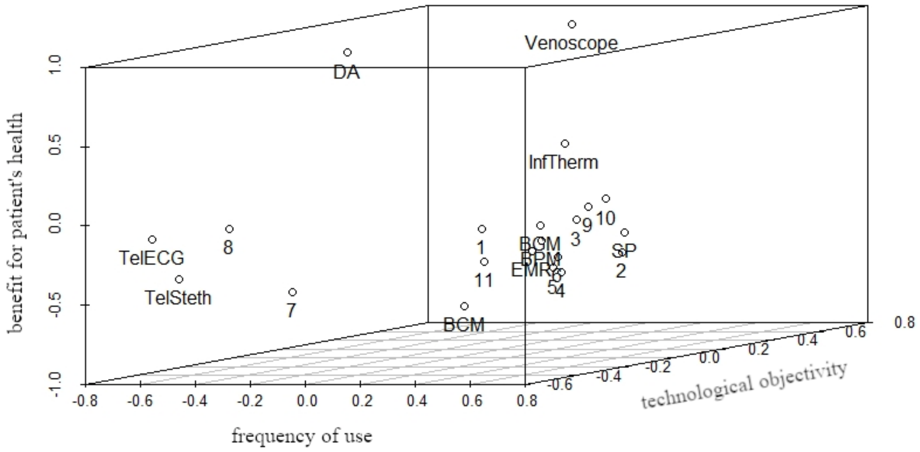
Considering Figure 3, we propose *technological objectivity* as an explanation for the second dimension. Except for the infrared thermometer, technologies sorted on the right, such as telemedical ECG or blood coagulation monitor, measure a patient’s more objective medical parameters, whereas technologies to the left, such as smartphones or EHR, are not used to directly assess a specific medical parameter, but rather to facilitate the communication and documentation of care. Participants are distributed more equally among the second dimension than among the first dimension, indicating that the perceived relevance associated with *technological objectivity* is a more subjective factor and that participants have a relatively ambivalent perspective on *technological objectivity*. Furthermore, an accumulation of most participants in the middle could indicate that *technological objectivity* might not directly increase the subjective relevance of technology use but that it depends on the specific context in which the technology is used. Overall, this dimension might represent the individual perspective of participants regarding the relevance of objective medical data.

Finally, we propose *benefit for a patient’s health status* as an explanation for our third dimension. Considering the technologies that appear at the bottom of Figure 4, for example, blood coagulation monitoring or telemedical stethoscope, participants might connect the characteristics of these technologies with the benefits they could have for a patient, especially in critical medical situations, as blood coagulation levels or auscultation sounds are strong indicators of a patient’s health. Seemingly distorting our interpretation of the third dimension, the arrangement of smartphones and EHRs toward the bottom of our configuration might be explained by the smartphone’s ability to connect the user to other medical professionals, especially in a medical emergency. Meanwhile, an EHR could be essential for GPs or MAs in interpreting or classifying specific symptoms of a patient with regard to their medical history and, therefore, essential in arriving at an appropriate diagnosis, especially in critical situations. Participants accumulate in the middle of the third dimension, indicating that the perceived benefit from a patient’s perspective can be seen relatively within a specific range of interpretation. Therefore, the third dimension of our configuration partially represents the professional identity of GPs and MAs and their motivation to use (digital) technologies arriving at a specific treatment objective to create a benefit for patients.

In accordance with the initial objective of our survey study, our results suggest three latent dimensions of subjective technology valuation by GPs and MAs. First, participants appear to assess the (potential) value of a technology by its extent of practical implementation, that is, the *hypothetical* and *actual frequency of use.* Second, the purpose of a specific technology appears to be another major reference for participants to decide whether a specific technology is relevant to their daily work routines or not. Here, participants differentiated between technologies primarily assessing objective medical parameters and those primarily used to communicate or document, that is, *technological objectivity.* Finally, the subjective valuation of technology appears to be dependent from the participants’ impression of whether it creates values from the perspective of patients and their treatment, that is, *benefit for a patient’s health status*. Figure 5 represents the final 3D configuration, including our interpretations.

To ensure that the dimensions of participants’ perceptions are valid, we discussed and reflected our findings from MDU with 17 physicians and MAs from our sample, as recommended by the literature [56,57], during a project-related workshop in November 2020. Participants confirmed our results from their professional perspective by emphasizing that our derived dimensions are plausible factors for their perception of technological relevance and effectiveness.

Figure 5. 3D configuration with interpreted dimensions (numbers represent participants, words represent different technologies). BCM: blood coagulation monitor; BGM: blood glucose monitor; BPM: blood pressure monitor; DA: digital appointment; EHR: electronic health records; InfTherm: infrared thermometer; SP: smartphone; TelECG: telemedical electrocardiogram; TelSteth: telemedical stethoscope.



Discussion

Overview

Reflecting the results from both the semistructured interviews and the survey study, we discussed the role of medical delegation for future primary care and the preconditions for a sustainable and effective implementation. Thereafter, we reasoned about specific aspects that are worth considering in the process of digitalizing primary care through mediated delegation and are needed to reflect possible (unintended) changes regarding the interaction between patients and medical professionals, that is, the subjectivity of medical data and the key role of MAs. Finally, we discussed possible challenges in the process of digitalizing primary care through mediated delegation and implications for theory and practice.

Mediated Delegation

As it is an essential prerequisite for multi-actor care processes, our data show that GPs attribute high degrees of competence and expertise to MAs when it comes to assessing a patient's health status and interpreting symptoms and parameters. Hence, MAs can pre-evaluate patient data and communicate it to the GP. In this process, both sides benefit from the integration of telemedical tools, such as digital measurement equipment and sensors, as the objective data they produce better inform the assessment of MAs. As a result, these tools expand the scope of what a GP can delegate and what an MA can actually do on site with the patient. As proposed by the diffusion of innovations model [70], the adoption of a technology also depends on the type of decision-making process. In particular, collective evaluations and decisions as they are made in multi-actor networks can facilitate higher adoption rates once required attributes, such as compatibility, are universally perceived. Here, compatibility plays a major role as the MA's competencies and working styles as well as the technology used need to fit the delegated task and requirements—both on an organizational and medical level—that come with it. Accordingly, the symbiosis of an MA gathering subjective and valuable impressions of the patient and technologies collecting objective data enables a more holistic image of the patient to be given to the GP, as our findings indicate that data of both subjective and objective nature are essential to arrive at a proper diagnosis and successful treatment protocol.

Demands and Requirements on Digital Technologies for Telemedicine

Another important factor that underlies the GPs' and MAs' perceptions of the use of technologies in direct patient care is the accuracy and instantaneousness of the applied sensors, the devices, and the resulting data. The participants stressed that the accuracy and quality of data is a major prerequisite for using digital technologies. For the evaluation of practicability and usefulness, GPs and MAs seem to consider the quality of medical data generated from a specific technology. Although usability might be relevant for technology use in general, participants rejected easier-to-use digital technologies because of their perceived lower precision. Furthermore, the objectiveness of technology is associated with participants' subjective perceptions of technological relevance, implying that the development of digital technologies for measuring medical parameters should address not only innovative ways to measure but should also ensure at least a constant quality of data compared with already elaborated technologies and procedures. Moreover, the consideration of legal frameworks or boundaries at an early stage in the development process might increase the adoption of innovative digital technologies because of the GPs' desire for cost-effectiveness, as the reimbursement of technology-mediated treatment by (social) health insurance companies seems to be an important factor for medical professionals when assessing a technology's usefulness [71]. In addition, financial support for the implementation of innovative technologies by public institutions might be a vital measure to counter the increasing workload and complexity of primary care. These findings align with current research on the success of eHealth interventions, identifying limited financial resources as a major obstacle [72].

Subjectivity of Medical Data

When it comes to the interpretation of medical data, our results suggest that GPs and MAs consider several subjective and environmental factors to be highly relevant, such as the state of a patient's home, medical history, or complex sensory impressions. Although common telemedical solutions such as video consultation systems enable practitioners to derive an initial picture of the patient based on what the camera captures and what the patient verbalizes, such technologies are still perceived as limited [73], particularly by hindering the continuance of care [74]. With digital technologies enabling the remote evaluation of a patient's health, the importance of combining or merging different technological characteristics becomes apparent, that is, the development of multichannel systems to provide not only quantified data but also a comprehensive image of a patient as well as platforms or opportunities for social interaction between the GP, MA, and patient. Consulting the diffusion of innovation model [70], it becomes apparent that a high level of complexity that comes with introducing a novel technology into multi-actor processes, paired with a lack of observability due to subjective assessments, highly individual routines, and tacit knowledge of involved actors, can further hamper the comprehensive adoption of telemedicine in multi-actor environments. Here, the complementary approach consisting of multi-actor care services and digital technologies facilitates the holistic assessment of the patient's health status and symptoms, as subjective impressions are augmented with digital measurements. As a result, the limitations of telemedicine are partially mitigated by the personal interaction of an MA with the patient, which in turn can lead to higher use intentions and adoption rates in the medical domain. Nevertheless, our findings show that a simple replacement of physical interaction between patient and MAs or GPs with digital interaction might lead to decreasing quality of care, because of a restriction of sensory information (eg, the state of a patient's environment, odors, or mobility) and the absence of physical contact to suggest caring [73]. At worst, a strategy of digitalizing primary care only from the perspective of economical effectiveness might lead to the discrimination of vulnerable groups, such as older, chronic, and immobile patients treated through ambulant primary care.

Telemedicine and the Key Role of MAs

The findings further indicate that the degree of responsibility and accountability with regard to actions performed by MAs when visiting a patient at home increases through the application of supportive telemedical tools. The data show that an MA, who is equipped with digital technologies such as telemedical stethoscopes or telemedical ECGs, is empowered to conduct a broad spectrum of diagnostic measures in a more autonomous and deliberate way. As a side effect, the presence of a supportive technology provides additional assurance to the MA as it augments the MA's subjective assessments with objective data, for instance, in the form of vital parameter measurements and visualizations. Furthermore, digital ICTs are able to reassure MAs by providing an enhanced amount of information about a patient, for example, mobile apps of EHRs or an enhanced remote interaction between the GP, MA, and patient by bridging spatial and temporal limitations [75]. Another interesting implication drawn from our results is that an MA potentially possesses high degrees of both medical and technical competence. Hence, in multi-actor treatment settings involving GPs, MAs, and patients, an MA who is skillful in using, understanding, and explaining digital technologies and their purpose is able to bridge gaps in technology competence and adoption behavior that occur on both the side of the care provider (ie, GP) and the side of the consumer (ie, patient).

Considering the major challenges that arise from the individual use of telemedicine, such as the insecurity of patients (and their relatives) to assist in a remote examination by manipulating technology [76], our findings suggest that technology-mediated delegation might be a suitable way to improve workflows of primary care physicians and address issues that arise from remote consultations. In general, the literature has shown that practitioners, as well as patients, have certain reservations when it comes to forming attitudes and use intentions toward digital health care technologies. On the professional side, empirical results indicate that some GPs think that a large portion of their patients, especially older patients, are not able to operate digital technologies, which hinders the effectiveness and progress of treatments [77]. On the patient side, low technology adoption rates can occur, inter alia, because of the desire to maintain a personal and direct relationship to their GP or the lack of familiarity with the respective technologies [71]. In many cases, patients do not see the benefit that digital technologies can have on their treatment and, thus, also on their health status and progression, because of their preuser status and unfamiliarity with innovative digital health care technologies [78,79]. Therefore, the integration of factors related to technology adoption in multi-actor settings, such as external perceptions of technological competence, might provide theoretical insights that could explain variations in the explanatory power of elaborated acceptance models with regard to health care technologies [47]. In this regard, our results indicate that the approach of combining multi-actor treatment settings with supportive digital technologies can bridge structural and perceptual gaps as well as a possible digital divide. As telemedicine becomes increasingly important for primary care, older patients may face disadvantages with regard to the delivery of vital care because of the underutilization of internet use and digital technologies [74,80]. Patients are no longer obliged to adopt technologies themselves and can stick with familiar treatment patterns while maintaining direct personal contact with the treating person. Especially for older patients or patients with chronic diseases lacking competence or interest in digital technologies [74,81], technology-mediated delegation might be an essential step to adopt innovative digital approaches to primary care. In addition, GPs are more likely to integrate technological tools into their working routines, as they trust their MAs to use them efficiently. Hence, the perceived lack of willingness and the ability to adopt digitalized treatments can be resolved.

Challenges Regarding Telemedicine Adoption

Furthermore, our results suggest that the adoption of digital technologies by GPs and MAs may be partially explained by the theoretical concept of bounded rationality. Considering the assumption that human decisions are not entirely based on a rational balancing of costs and benefits, but on heuristics and cognitive simplifications as well [82], the tendency to prefer familiar technologies, called the status quo bias, is known to have a powerful effect and could explain user resistance [83]. Most participants in our studies found the technologies that they had already learned how to integrate in their daily work routine especially relevant and, therefore, potentially assessed the subjective relevance of unfamiliar digital technologies, such as telemedical applications, from the perspective of their status quo. While anchoring their subjective perception of usefulness and relevance to familiar technologies [84], technologies containing innovative characteristics that are dissimilar to elaborated ones are possibly rejected. Thus, considering that the evaluation and adoption of technologies partially depend on heuristics in decision-making that are often applied subconsciously and are thus difficult to externalize and become aware of [85], additional barriers occur that potentially hamper the comprehensive adoption of telemedicine in multi-actor care. Therefore, to increase the attraction of innovative technologies in primary care, providers might consider the importance of the integrability of their products into the existing routines and communicative structures of their customers. The possibility of using digital technologies without an obligation to eventually buy them might hold an opportunity for customers, that is, GPs and the staff of their practices, to experimentally incorporate a new technology to overcome the status quo bias and the anchoring effects to which they are subject. On the contrary, as the aforementioned heuristics serve the goal of making fast and efficient decisions and in the light of lacking information regarding the application and performance of a given technology [86,87], it should be further taken into account that in some specific cases, digitally complementing treatments in multi-actor care do not represent a feasible and effective way of making things better, but instead introduces new forms of overhead and uncertainty.

Theoretical and Practical Implications

Through our empirical findings, we provide insights into digital technologies and their potential for multi-actor delegation processes in primary care in relation to research, medical practice, and the development and design of technology. For theoretical reasoning, our paper points out several factors that are relevant for expanding the understanding of technology use by medical professionals. Within the context of multi-actor approaches to delegation in primary care, the results suggest that MAs facilitate the use of and access to health care technology for patients. There, MAs might even be able to compensate for a possible lack of technological competence or skills and for the resistance of patients bypassing obstacles related to usability [88]. In this regard, our results may be used to expand or contextualize theories concerning technological adoption used for telemedicine [47], such as the Technology Acceptance Model or the unified theory of acceptance and use of technology. Technology adoption and use appear to be reciprocal processes involving external perceptions of patients and their hypothetical reactions to the use of digital technologies by medical professionals, which might be considered to enrich theoretical models. In addition, we explored an anchoring effect [84] concerning the adoption of unfamiliar technologies by GPs and MAs, indicating that the adoption of digital technologies in health care might comprise cognitive biases that need to be further differentiated and operationalized to improve the predictive power of elaborated theoretical models.

From a practical point of view, our findings indicate a strong potential and benefit of using digital technologies, especially those used for telemedical examinations and video consultation for delegation processes in ambulant primary care. This holds not only for GP-MA interaction but also between GPs and nurses or registered nurses and nursing assistants. Owing to a shift in the organization of health care that can be noticed in North and Middle European, or North American countries, the relevance of team-based primary care increases [89-91]. Our study suggests that technology-mediated delegation and care processes might be a suitable way to ensure comprehensive personal care and effectively deal with the challenges of time pressure, increasing case complexity, and shortage of medical professionals. In this regard, our results are valuable for the strategic alignment of health care providers, health insurance companies, companies developing telemedicine applications, or politics. Moreover, the potential discharge of health care systems' resources through differentiated and needs-based care has become increasingly important with respect to the ongoing COVID-19 pandemic. The ability to guarantee adequate care for patients at home through MAs utilizing telemedicine might increase time-efficient treatment and accurate monitoring of patients while decreasing inequalities concerning personal technological competence or affinity between younger and older patients. Within the multi-actor processes of primary care, digital technologies can enrich the subjective assessment of a patient by optimizing the objective measurements of medical data and by providing a more effective way to communicate and document. By comprehending the subjective interaction between MAs and patients, specific prerequisites for innovative digital technology use might change or be omitted because of the medical and technological competence of MAs. Nevertheless, our study suggests that the accuracy and reliability of the technology remain an important factor for medical professionals.

Limitations

The results of our paper were bound to relatively small sample sizes. Although the explorative nature of our research objective matched our chosen methods of data collection and analysis, a potential bias might emerge from the fact that both studies were drawn from the same sample. In addition, data from our survey study represent only a part, but not the entire sample of our interview study, resulting in limited generalizability within the respective sample. Although GPs and MAs ensured the validity of our results from our survey study, our findings might partially reflect the tendency to confirm existing knowledge (from our interview study), rather than disprove them, known as confirmation bias [92]. Therefore, our selected method for data analysis (MDU) generally implies the risk of subjective interpretation and needs to be discussed in future research. Furthermore, as all participants in our studies were associated with a regional project on the digitalization of health care technologies for delegation processes in ambulant medical care, participants might represent opinions and perceptions that tend to be more optimistic and interested in the ongoing process of digitalization. In addition, we concentrated on a sample related to specific characteristics, that is, GPs and MAs in rural German areas. To extend the validity of our findings, a quantitative approach with a large sample size might provide insights into the generalizability of our findings and optimize the representativeness of our MDU model. As our study provides insights into the process of technology adoption through incorporation and shows the importance of practical experience for technology adoption, an interventional pre-post study might further enrich our findings.

Conclusions

Our study explores the potential of using digital technologies in primary care delegation processes. Interviews with GPs and MAs revealed the complex situational role of technology within these delegation processes. Although the results suggest that the importance of ICT is increasing because of its ability to remove spatial and temporal limitations, telemedical solutions appear to be promising, as they enable video consultation or automated transfer of medical data. In addition, telemedical solutions have the potential to facilitate direct patient treatment by merging medical and social competence to overcome demographic and structural changes, as well as to overcome deficits in patients' technological competence. Therefore, digital technologies assist in finding innovative, case-sensitive, and cost-effective methods of treatment in primary care. Nonetheless, our study revealed the contextual nature of technology use in primary care. The adoption and implementation of technology underlie reciprocal processes involving different attitudes and perceptions within multi-actor settings. Furthermore, the results suggest that these attitudes and perceptions might be biased because of the underlying needs for action that are unique to medical treatments. Consequently, our study provides a foundation for further investigation of relational characteristics within multi-actor settings in primary care. Finally, practical suggestions are made to improve the development and distribution of innovative technologies for medical delegation processes.

Acknowledgments

This study was conducted within the research project NaePas and supported by a grant from the district administration of Altenkirchen (Rhineland-Palatine, Germany).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guidelines.
[DOCX File , 15 KB - jmir_v23i8e28151_app1.docx]

References

1. Demiris G, Hensel BK. Technologies for an aging society: a systematic review of “Smart Home” applications. Yearb Med Inform 2018 Mar 07;17(01):33-40. [doi: 10.1055/s-0038-1638580]
2. Adarkwah CC, Schwaffertz A, Labenz J, Becker A, Hirsch O. Burnout and work satisfaction in general practitioners practicing in rural areas: results from the HaMEdSi study. Psychol Res Behav Manag 2018 Oct;Volume 11:483-494. [doi: 10.2147/prbm.s179503]
3. Yang J. Potential urban-to-rural physician migration: The limited role of financial incentives. Can J Rur Med 2003;8(2):101-106 [FREE Full text]
4. Adarkwah CC, Schwaffertz A, Labenz J, Becker A, Hirsch O. [Assessment of the occupational perspectives of general practitioners in a rural area. Results from the study HaMedSi (Hausärzte [GPs] for Medical education in Siegen-Wittgenstein)]. MMW Fortschr Med 2019 Oct;161(Suppl 6):9-14. [doi: 10.1007/s15006-019-0919-4] [Medline: 31587169]
5. Politzer RM, Yoon J, Shi L, Hughes RG, Regan J, Gaston MH. Inequality in America: the contribution of health centers in reducing and eliminating disparities in access to care. Med Care Res Rev 2001 Jun;58(2):234-248. [doi: 10.1177/107755870105800205] [Medline: 11398647]
6. Wilson N, Couper I, De Vries E, Reid S, Fish T, Marais B. A critical review of interventions to redress the inequitable distribution of healthcare professionals to rural and remote areas. Rur Rem Health 2009;9(2):1060 [FREE Full text] [Medline: 19530891]
7. 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]
8. Reeves JJ, Ayers JW, Longhurst CA. Telehealth in the COVID-19 Era: A balancing act to avoid harm. J Med Internet Res 2021 Feb 01;23(2):e24785 [FREE Full text] [doi: 10.2196/24785] [Medline: 33477104]
9. Valentino LA, Skinner MW, Pipe SW. The role of telemedicine in the delivery of health care in the COVID-19 pandemic. Haemophilia 2020 Sep;26(5):230-231 [FREE Full text] [doi: 10.1111/hae.14044] [Medline: 32397000]
10. Jalali M, Landman A, Gordon W. Telemedicine, privacy, and information security in the age of COVID-19. J Am Med Inform Assoc 2021 Mar 01;28(3):671-672 [FREE Full text] [doi: 10.1093/jamia/ocaa310] [Medline: 33325533]
11. Williams CM, Chaturvedi R, Chakravarthy K. Cybersecurity risks in a pandemic. J Med Internet Res 2020 Sep 17;22(9):e23692 [FREE Full text] [doi: 10.2196/23692] [Medline: 32897869]

12. Sheridan B, Chien A, Peters A, Rosenthal M, Brooks J, Singer S. Team-based primary care: The medical assistant perspective. Health Care Manage Rev 2018;43(2):115-125. [doi: 10.1097/HMR.0000000000000136] [Medline: 27849646]
13. Anderson A, Proudfoot J, Harris M. Medical assistants: a primary care workforce solution? Aust Fam Physician 2009 Aug;38(8):623-626 [FREE Full text] [Medline: 19893785]
14. Chapman SA, Blash LK. New roles for medical assistants in innovative primary care practices. Health Serv Res 2017 Feb;52 Suppl 1:383-406 [FREE Full text] [doi: 10.1111/1475-6773.12602] [Medline: 27859097]
15. Günther HJ, Bader C, Erlenberg RM, Hagl C, Schirrmacher B, Schuster A. [From AGnES to PA - medical assistant professions in Germany: who still keeps the track?]. MMW Fortschr Med 2019 Dec;161(Suppl 7):21-30. [doi: 10.1007/s15006-019-1175-3] [Medline: 31828670]
16. Gray C, Harrison M, Hung D. Medical assistants as flow managers in primary care: challenges and recommendations. J Healthc Manag 2016;61(3):181-191. [Medline: 27356444]
17. Echelard J, Méthot F, Nguyen H, Pomey M. Medical student training in eHealth: scoping review. JMIR Med Educ 2020 Sep 11;6(2):e20027 [FREE Full text] [doi: 10.2196/20027] [Medline: 32915154]
18. Almathami HK, Win KT, Vlahu-Gjorgievska E. Barriers and facilitators that influence telemedicine-based, real-time, online consultation at patients' homes: systematic literature review. J Med Internet Res 2020 Feb 20;22(2):e16407. [doi: 10.2196/16407]
19. Vesnic-Alujevic L, Breitegger M, Guimarães Pereira Â. 'Do-It-Yourself' Healthcare? Quality of health and healthcare through wearable sensors. Sci Eng Ethics 2018 Jun;24(3):887-904 [FREE Full text] [doi: 10.1007/s11948-016-9771-4] [Medline: 27029478]
20. Aamodt IT, Lycholip E, Celutkiene J, Strömberg A, Atar D, Falk RS, et al. Health care professionals' perceptions of home telemonitoring in heart failure care: cross-sectional survey. J Med Internet Res 2019 Feb 06;21(2):e10362 [FREE Full text] [doi: 10.2196/10362] [Medline: 30724744]
21. Yin Y, Zeng Y, Chen X, Fan Y. The internet of things in healthcare: an overview. J Ind Inf Integ 2016 Mar;1:3-13. [doi: 10.1016/j.jii.2016.03.004]
22. Habibzadeh H, Dinesh K, Shishvan O, Boggio-Dandry A, Sharma G, Soyata T. A Survey of Healthcare Internet-of-Things (HIoT): A clinical perspective. IEEE Internet Things J 2020 Jan;7(1):53-71 [FREE Full text] [doi: 10.1109/jiot.2019.2946359] [Medline: 33748312]
23. Appireddy R, Khan S, Leaver C, Martin C, Jin A, Durafourt BA, et al. Home virtual visits for outpatient follow-up stroke care: cross-sectional study. J Med Internet Res 2019 Oct 07;21(10):e13734 [FREE Full text] [doi: 10.2196/13734] [Medline: 31593536]
24. Seuren LM, Wherton J, Greenhalgh T, Cameron D, A'Court C, Shaw SE. Physical examinations via video for patients with heart failure: qualitative study using conversation analysis. J Med Internet Res 2020 Feb 20;22(2):e16694 [FREE Full text] [doi: 10.2196/16694] [Medline: 32130133]
25. Mayakul T, Ayuthaya S. A digital prescription refill system based on healthcare standard in Thailand. Int J Appl Biomed Eng 2018;11(1):8 [FREE Full text]
26. Nazia S, Ekta S. Online appointment scheduling system for hospitals - an analytical study. Int J Innov Eng Technol 2014;4(1):21-27 [FREE Full text]
27. Ahmad F, Hudak PL, Bercovitz K, Hollenberg E, Levinson W. Are physicians ready for patients with internet-based health information? J Med Internet Res 2006 Sep 29;8(3):e22 [FREE Full text] [doi: 10.2196/jmir.8.3.e22] [Medline: 17032638]
28. Peine A, Paffenholz P, Martin L, Dohmen S, Marx G, Loosen SH. Telemedicine in Germany during the COVID-19 pandemic: multi-professional national survey. J Med Internet Res 2020 Aug 05;22(8):e19745 [FREE Full text] [doi: 10.2196/19745] [Medline: 32568724]
29. Salisbury C, Murphy M, Duncan P. The impact of digital-first consultations on workload in general practice: modeling study. J Med Internet Res 2020 Jun 16;22(6):e18203 [FREE Full text] [doi: 10.2196/18203] [Medline: 32543441]
30. Yeow A, Goh KH. Work harder or work smarter? Information technology and resource allocation in healthcare processes. MIS Q 2015 Apr 4;39(4):763-785. [doi: 10.25300/misq/2015/39.4.2]
31. Mueller M, Knop M, Niehaves B, Adarkwah CC. Investigating the acceptance of video consultation by patients in rural primary care: empirical comparison of preusers and actual users. JMIR Med Inform 2020 Oct 22;8(10):e20813 [FREE Full text] [doi: 10.2196/20813] [Medline: 32969339]
32. Mackey H, Nancarrow S. Assistant practitioners: Issues of accountability, delegation and competence. Int J Ther Rehab 2005 Aug;12(8):331-338. [doi: 10.12968/ijtr.2005.12.8.19535]
33. Riisgaard H, Søndergaard J, Munch M, Le J, Ledderer L, Pedersen L, et al. Associations between degrees of task delegation and job satisfaction of general practitioners and their staff: a cross-sectional study. BMC Health Serv Res 2017 Jan 17;17(44):1-9 [FREE Full text] [doi: 10.1186/s12913-017-1984-y] [Medline: 28095846]
34. Sibbald B, McDonald R, Roland M. Shifting care from hospitals to the community: a review of the evidence on quality and efficiency. J Health Serv Res Policy 2007 Apr 24;12(2):110-117. [doi: 10.1258/135581907780279611] [Medline: 17407662]
35. Laurant M, Reeves D, Hermens R, Braspenning J, Grol R, Sibbald B. Substitution of doctors by nurses in primary care. Cochrane Database Syst Rev 2005 Apr 18(2):CD001271. [doi: 10.1002/14651858.CD001271.pub2] [Medline: 15846614]

36. Mundinger M, Kane R, Lenz E, Totten A, Tsai W, Cleary P, et al. Primary care outcomes in patients treated by nurse practitioners or physicians: a randomized trial. J Am Med Assoc 2000 Jan 05;283(1):59-68. [doi: 10.1001/jama.283.1.59] [Medline: 10632281]
37. Naughton D, Adelman A, Bricker P, Miller-Day M, Gabbay R. Envisioning new roles for medical assistants: strategies from patient-centered medical homes. Fam Pract Manag 2013;20(2):7-12 [FREE Full text] [Medline: 23547608]
38. Harper P, Van Riper K, Ramer T, Slattengren A, Adam P, Smithson A, et al. Team-based care: an expanded medical assistant role - enhanced rooming and visit assistance. J Interprof Care 2018 Nov 02:1-7. [doi: 10.1080/13561820.2018.1538107] [Medline: 30388911]
39. Rokicki-Parashar J, Phadke A, Brown-Johnson C, Jee O, Sattler A, Torres E, et al. Transforming interprofessional roles during virtual health care: the evolving role of the medical assistant, in relationship to national health profession competency standards. J Prim Care Community Health 2021;12:21501327211004285 [FREE Full text] [doi: 10.1177/21501327211004285] [Medline: 33764223]
40. Côté N, Freeman A, Jean E, Denis JL. New understanding of primary health care nurse practitioner role optimisation: the dynamic relationship between the context and work meaning. BMC Health Serv Res 2019 Nov 21;19(1):1-10 [FREE Full text] [doi: 10.1186/s12913-019-4731-8] [Medline: 31752860]
41. Swanson M, Wong S, Martin-Misener R, Browne A. The role of registered nurses in primary care and public health collaboration: a scoping review. Nurs Open 2020 Jul;7(4):1197-1207 [FREE Full text] [doi: 10.1002/nop2.496] [Medline: 32587740]
42. Mohr D, Young G, Meterko M, Stolzmann K, White B. Job satisfaction of primary care team members and quality of care. Am J Med Qual 2011;26(1):18-25. [doi: 10.1177/1062860610373378] [Medline: 20935270]
43. Szecsenyi J, Goetz K, Campbell S, Broge B, Reuschenbach B, Wensing M. Is the job satisfaction of primary care team members associated with patient satisfaction? BMJ Qual Saf 2011 Jun 24;20(6):508-514 [FREE Full text] [doi: 10.1136/bmjqs.2009.038166] [Medline: 21262790]
44. Illario M, Coscioni E, De Luca V, Cataldi M, Postiglione A, Iaccarino G. Editorial: Multi-actor collaboration in healthcare to address the emerging health needs of an aging population. Transl Med UniSa 2019;19:1-4 [FREE Full text] [Medline: 31360660]
45. Fürstenau D, Auschra C, Klein S, Gersch M. A process perspective on platform design and management: evidence from a digital platform in health care. Electron Markets 2018 Dec 13;29(4):581-596. [doi: 10.1007/s12525-018-0323-4]
46. Bleyel C, Hoffmann M, Wensing M, Hartmann M, Friederich H, Haun M. Patients' perspective on mental health specialist video consultations in primary care: qualitative preimplementation study of anticipated benefits and barriers. J Med Internet Res 2020 Apr 20;22(4):e17330 [FREE Full text] [doi: 10.2196/17330] [Medline: 32310139]
47. Harst L, Lantzsch H, Scheibe M. Theories predicting end-user acceptance of telemedicine use: systematic review. J Med Internet Res 2019 May 21;21(5):e13117 [FREE Full text] [doi: 10.2196/13117] [Medline: 31115340]
48. Lakhe A, Sodhi I, Warrier J, Sinha V. Development of digital stethoscope for telemedicine. J Med Eng Technol 2016 Jan 05;40(1):20-24. [doi: 10.3109/03091902.2015.1116633] [Medline: 26728637]
49. Wickramasinghe N, John B, George J, Vogel D. Achieving value-based care in chronic disease management: intervention study. JMIR Diabetes 2019 May 03;4(2):e10368 [FREE Full text] [doi: 10.2196/10368] [Medline: 31066699]
50. Agarwal R, Gao G, DesRoches C, Jha AK. Research Commentary—The digital transformation of healthcare: current status and the road ahead. Inf Syst Res 2010 Dec;21(4):796-809. [doi: 10.1287/isre.1100.0327]
51. Venkatesh V, Brown SA, Bala H. Bridging the Qualitative-Quantitative Divide: Guidelines for conducting mixed methods research in information systems. MIS Q 2013 Jan 1;37(1):21-54. [doi: 10.25300/misq/2013/37.1.02]
52. Venkatesh V, Brown S, Sullivan Y. Guidelines for conducting mixed-methods research: an extension and illustration. J Asso Inf Syst 2016 Aug;17(7):435-494. [doi: 10.17705/1jais.00433]
53. Marshall MN. Sampling for qualitative research. Fam Pract 1996 Dec;13(6):522-525. [doi: 10.1093/fampra/13.6.522] [Medline: 9023528]
54. Strauss A, Corbin J. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. Thousand Oaks (CA): Sage Publications Inc; 1998.
55. Childs RA, Ram A, Xu Y. Combining dual scaling with semi-structured interviews to interpret rating differences. Pract Assess Res Eval 2009;14(1):1-9 [FREE Full text] [doi: 10.7275/wtdw-1c03]
56. Raghavan C, Harkness S, Super C. Parental ethnotheories in the context of immigration: Asian Indian immigrant and Euro-American mothers and daughters in an American town. J Cross Cult Psychol 2010 Jul 20;41(4):617-632. [doi: 10.1177/0022022110362629]
57. Matook S, Vessey I. Types of business-to-business e-marketplaces: The role of a theory-based, domain-specific model. J Electron Comm Res 2008;9(4):60-279 [FREE Full text]
58. Goode S, Gregor S. Rethinking organisational size in IS research: meaning, measurement and redevelopment. Eur J Inf Syst 2017 Dec 19;18(1):4-25. [doi: 10.1057/ejis.2009.2]
59. Posey C, Roberts TL, Lowry PB, Bennett RJ, Courtney JF. Insiders’ protection of organizational information assets: Development of a systematics-based taxonomy and theory of diversity for protection-motivated behaviors. MIS Q 2013 Apr 4;37(4):1189-1210. [doi: 10.25300/misq/2013/37.4.09]

60. Venkatraman S, Cheung C, Lee Z, Davis F, Venkatesh V. The “Darth” side of technology use: an inductively derived typology of cyberdeviance. J Manag Inf Syst 2018 Dec 17;35(4):1060-1091. [doi: 10.1080/07421222.2018.1523531]
61. Rodgers JL. Matrix and stimulus sample sizes in the weighted mds model: Empirical metric recovery functions. Appl Psychol Measur 2016 Jul 26;15(1):71-77. [doi: 10.1177/014662169101500107]
62. Borg I, Groenen P. Modern Multidimensional Scaling: Theory and Applications. New York (NY): Springer Science & Business Media; 2005.
63. Borg I, Groenen P, Mair P. Applied Multidimensional Scaling. Heidelberg: Springer; 2013.
64. R: A language environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2020. URL: https://www.R-project.org [accessed 2019-02-01]
65. Leeuw JD, Mair P. Multidimensional scaling using majorization: SMACOF in R. J Stat Soft 2009;31(3):1-30. [doi: 10.18637/jss.v031.i03]
66. Martins AA, Cardoso MG, Pinto IM. Multidimensional unfolding and the use of alternative input scales. J Target Meas Anal Mark 2009 Jun 29;17(3):195-204. [doi: 10.1057/jt.2009.10]
67. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425-478. [doi: 10.2307/30036540]
68. Busing FM, Groenen PJ, Heiser WJ. Avoiding degeneracy in multidimensional unfolding by penalizing on the coefficient of variation. Psychometrika 2005 Apr 10;70(1):71-98. [doi: 10.1007/s11336-001-0908-1]
69. Ligges U, Mächler M. scatterplot3d - An R package for visualizing multivariate data. J Stat Soft 2003;8(11):1-20. [doi: 10.18637/jss.v008.i11]
70. Rogers E. Diffusion of Innovations. New York: Free Press; 1995:1-518.
71. Thies K, Anderson D, Cramer B. Lack of adoption of a mobile app to support patient self-management of diabetes and hypertension in a federally qualified health center: interview analysis of staff and patients in a failed randomized trial. JMIR Hum Factors 2017 Oct 03;4(4):e24 [FREE Full text] [doi: 10.2196/humanfactors.7709] [Medline: 28974481]
72. Granja C, Janssen W, Johansen MA. Factors determining the success and failure of ehealth interventions: systematic review of the literature. J Med Internet Res 2018 May 01;20(5):e10235 [FREE Full text] [doi: 10.2196/10235] [Medline: 29716883]
73. Knop M, Mueller M, Freude H, Reßing C, Niehaves B. Perceived limitations of telemedicine from a phenomenological perspective. In: Proceedings of the 33rd BLED eConference. 2020 Presented at: 33rd BLED eConference; June 28 - July 1, 2020; Maribor, Slovenia URL: https://aisel.aisnet.org/bled2020/34/ [doi: 10.18690/978-961-286-362-3.9]
74. Salisbury C, Quigley A, Hex N, Aznar C. Private video consultation services and the future of primary care. J Med Internet Res 2020 Oct 01;22(10):e19415 [FREE Full text] [doi: 10.2196/19415] [Medline: 32812887]
75. Kvedar J, Coye MJ, Everett W. Connected health: a review of technologies and strategies to improve patient care with telemedicine and telehealth. Health Aff (Millwood) 2014 Feb;33(2):194-199. [doi: 10.1377/hlthaff.2013.0992] [Medline: 24493760]
76. Shaw SE, Seuren LM, Wherton J, Cameron D, A'Court C, Vijayaraghavan S, et al. Video consultations between patients and clinicians in diabetes, cancer, and heart failure services: linguistic ethnographic study of video-mediated interaction. J Med Internet Res 2020 May 11;22(5):e18378 [FREE Full text] [doi: 10.2196/18378] [Medline: 32391799]
77. Mueller M, Knop M, Reßing C, Freude H, Oschinsky F, Klein H, et al. Constituting factors of a digitally influenced relationship between patients and primary care physicians in rural areas. In: Proceedings of the 53rd Hawaii International Conference on System Sciences (HICSS). 2020 Presented at: 53rd Hawaii International Conference on System Sciences (HICSS); January 7-10, 2020; Hawaii, USA. [doi: 10.24251/hicss.2020.447]
78. Bagozzi R. The legacy of the technology acceptance model and a proposal for a paradigm shift. J Asso Inf Syst 2007 Apr;8(4):244-254. [doi: 10.17705/1jais.00122]
79. Karsh B, Weinger MB, Abbott PA, Wears RL. Health information technology: fallacies and sober realities. J Am Med Inform Assoc 2010 Nov 01;17(6):617-623 [FREE Full text] [doi: 10.1136/jamia.2010.005637] [Medline: 20962121]
80. Niehaves B, Plattfaut R. Internet adoption by the elderly: employing IS technology acceptance theories for understanding the age-related digital divide. Eur J Inf Syst 2017 Dec 19;23(6):708-726. [doi: 10.1057/ejis.2013.19]
81. McGrail KM, Ahuja MA, Leaver CA. Virtual visits and patient-centered care: results of a patient survey and observational study. J Med Internet Res 2017 May 26;19(5):e177 [FREE Full text] [doi: 10.2196/jmir.7374] [Medline: 28550006]
82. Kahneman D. Maps of bounded rationality: Psychology for behavioral economics. Am Econ Rev 2003 Nov 01;93(5):1449-1475. [doi: 10.1257/000282803322655392]
83. Kim H, Kankanhalli A. Investigating user resistance to information systems implementation: a status quo bias perspective. MIS Q 2009;33(3):567. [doi: 10.2307/20650309]
84. Lee K, Joshi K. Examining the use of status quo bias perspective in IS research: need for re-conceptualizing and incorporating biases. Info Systems J 2016 Jul 29;27(6):733-752. [doi: 10.1111/isj.12118]
85. Schirrmeister E, Göhring A, Warnke P. Psychological biases and heuristics in the context of foresight and scenario processes. Fut Foresight Sci 2020 Feb 03;2(2):e31. [doi: 10.1002/ffo2.31]
86. 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]

87. Simon HA. Invariants of human behavior. Annu Rev Psychol 1990 Jan;41(1):1-19. [doi: 10.1146/annurev.ps.41.020190.000245] [Medline: 18331187]
88. Samhan B. Patients' resistance towards health information technology: a perspective of the dual factor model of IT usage. In: Proceedings of the 50th Hawaii International Conference on System Sciences Hawaii. 2017 Presented at: 50th Hawaii International Conference on System Sciences (HICSS); January 4-7, 2017; Hawaii, USA p. 3401-3410. [doi: 10.24251/hicss.2017.412]
89. Wagner EH, Flinter M, Hsu C, Cromp D, Austin BT, Etz R, et al. Effective team-based primary care: observations from innovative practices. BMC Fam Pract 2017 Feb 02;18(1):13 [FREE Full text] [doi: 10.1186/s12875-017-0590-8] [Medline: 28148227]
90. Mitchell JD, Haag JD, Klavetter E, Beldo R, Shah ND, Baumbach LJ, et al. Development and implementation of a team-based, primary care delivery model: challenges and opportunities. Mayo Clin Proc 2019 Jul;94(7):1298-1303. [doi: 10.1016/j.mayocp.2019.01.038] [Medline: 31272572]
91. Rudebeck CE. Relationship based care - how general practice developed and why it is undermined within contemporary healthcare systems. Scand J Prim Health Care 2019 Sep 12;37(3):335-344 [FREE Full text] [doi: 10.1080/02813432.2019.1639909] [Medline: 31299870]
92. Johnson RB, Onwuegbuzie AJ. Mixed methods research: a research paradigm whose time has come. Edu Res 2016 Jul 01;33(7):14-26. [doi: 10.3102/0013189X033007014]

Abbreviations

ECG: electrocardiogram
EHR: electronic health record
GP: general practitioner
ICT: information and communication technology
MA: medical assistant
MDS: multidimensional scaling
MDU: multidimensional unfolding

Edited by R Kukafka; submitted 23.02.21; peer-reviewed by J Glass, J Parsons, J Hennrich, JF Echelard, M Gropalis; comments to author 05.05.21; revised version received 24.06.21; accepted 05.07.21; published 26.08.21.

Please cite as:
Knop M, Mueller M, Niehaves B
Investigating the Use of Telemedicine for Digitally Mediated Delegation in Team-Based Primary Care: Mixed Methods Study
J Med Internet Res 2021;23(8):e28151
URL: https://www.jmir.org/2021/8/e28151
doi:10.2196/28151
PMID:34435959

©Michael Knop, Marius Mueller, Bjoern Niehaves. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Review

Telehealth Interventions and Outcomes Across Rural Communities in the United States: Narrative Review

Michael Butzner[1*], DrPH; Yendelela Cuffee[1,2*], PhD

[1]Department of Public Health Sciences, College of Medicine, Pennsylvania State University, Hershey, PA, United States

[2]Program in Epidemiology, College of Health Sciences, University of Delaware, Newark, DE, United States

[*]all authors contributed equally

Corresponding Author:
Michael Butzner, DrPH
Department of Public Health Sciences
College of Medicine
Pennsylvania State University
90 Hope Drive
Suite 2200
Hershey, PA, 17033
United States
Phone: 1 4109678635
Email: mab1468@psu.edu

Abstract

Background: In rural communities, there are gaps in describing the design and effectiveness of technology interventions for treating diseases and addressing determinants of health.

Objective: The aim of this study is to evaluate literature on current applications, therapeutic areas, and outcomes of telehealth interventions in rural communities in the United States.

Methods: A narrative review of studies published on PubMed from January 2017 to December 2020 was conducted. Key search terms included telehealth, telemedicine, rural, and outcomes.

Results: Among 15 included studies, 9 studies analyzed telehealth interventions in patients, 3 in health care professionals, and 3 in both patients and health care professionals. The included studies reported positive outcomes and experiences of telehealth use in rural populations including acceptability and increased satisfaction; they also noted that technology is convenient and efficient. Other notable benefits included decreased direct and indirect costs to the patient (travel cost and time) and health care service provider (staffing), lower onsite health care resource utilization, improved physician recruitment and retention, improved access to care, and increased education and training of patients and health care professionals.

Conclusions: Telehealth models were associated with positive outcomes for patients and health care professionals, suggesting these models are feasible and can be effective. Future telehealth interventions and studies examining these programs are warranted, especially in rural communities, and future research should evaluate the impact of increased telehealth use as a result of the COVID-19 pandemic.

(J Med Internet Res 2021;23(8):e29575) doi:10.2196/29575

KEYWORDS

telehealth; telemedicine; rural health; health outcomes; social determinants of health; eHealth; health care accessibility

Introduction

Health care is increasingly becoming a technology-driven environment. Telehealth is a remote health care service delivery method and allows for real-time communication between a patient and health care provider [1]. Telehealth is an alternative model of health care service delivery; specifically, it provides opportunities to expand treatment access and reduce barriers to care in underserved and rural areas. Telehealth has been used to promote healthy behaviors and condition management, and there have been promising effects observed, including increasing patient participation and satisfaction and reducing rates of chronic illnesses. In one study, ter Huurne et al [2] conducted a web-based technology study to improve health care service for individuals with chronic eating disorders. In their study, which used a web-based treatment program to improve health care service, 54% of participants completed all program tasks and assignments, and the program significantly improved BMI, body dissatisfaction, quality of life, and physical and mental health [2].

Telehealth is a convenient approach for patients to access health care in the comfort of their own home. Kruse et al [3] conducted a literature review to examine the association of telehealth and patient satisfaction in terms of efficiency and effectiveness. The review concluded that telehealth decreases travel time, improves communication with providers, increases access to care, increases self-awareness, and empowers patients to manage their chronic conditions [3]. From a health care system and provider perspective, benefits include decreased missed appointments, decreased wait times, decreased readmissions, improved medication adherence, and improved quality and timeliness of patient care; in addition, telehealth is a good modality for education [3]. Physician shortage and burnout are common and significant issues in rural areas that can be alleviated by telehealth. Ward et al [4] found that 75% of family physicians in rural areas were covering local emergency departments (ED) as a condition of their practice or hospital privilege. Mandating these conditions discourages physicians from practicing in rural areas; however, the study concluded that telemedicine may help improve the chronic rural workforce shortage by improving physician recruitment and retention [4].

Residents of rural communities across the United States include some of the most vulnerable populations, including individuals with low socioeconomic status, Indigenous communities, children and older adults, and individuals with disabilities [1]. People living in rural communities have limited access to health care, travel long distances to receive care, and/or delay care until after they have a health emergency. Limited access to health care can result in poor health outcomes and is a social and economic burden for both the patient and the health care system. The cost associated with traveling for medical care places an additional burden on the patient, including incurring additional costs for traveling to visits, lost work hours, lower productivity, and increased costs associated with caregiver or childcare support [1]. Telehealth extends the reach of health services and provides the opportunity to reduce barriers to care in rural communities.

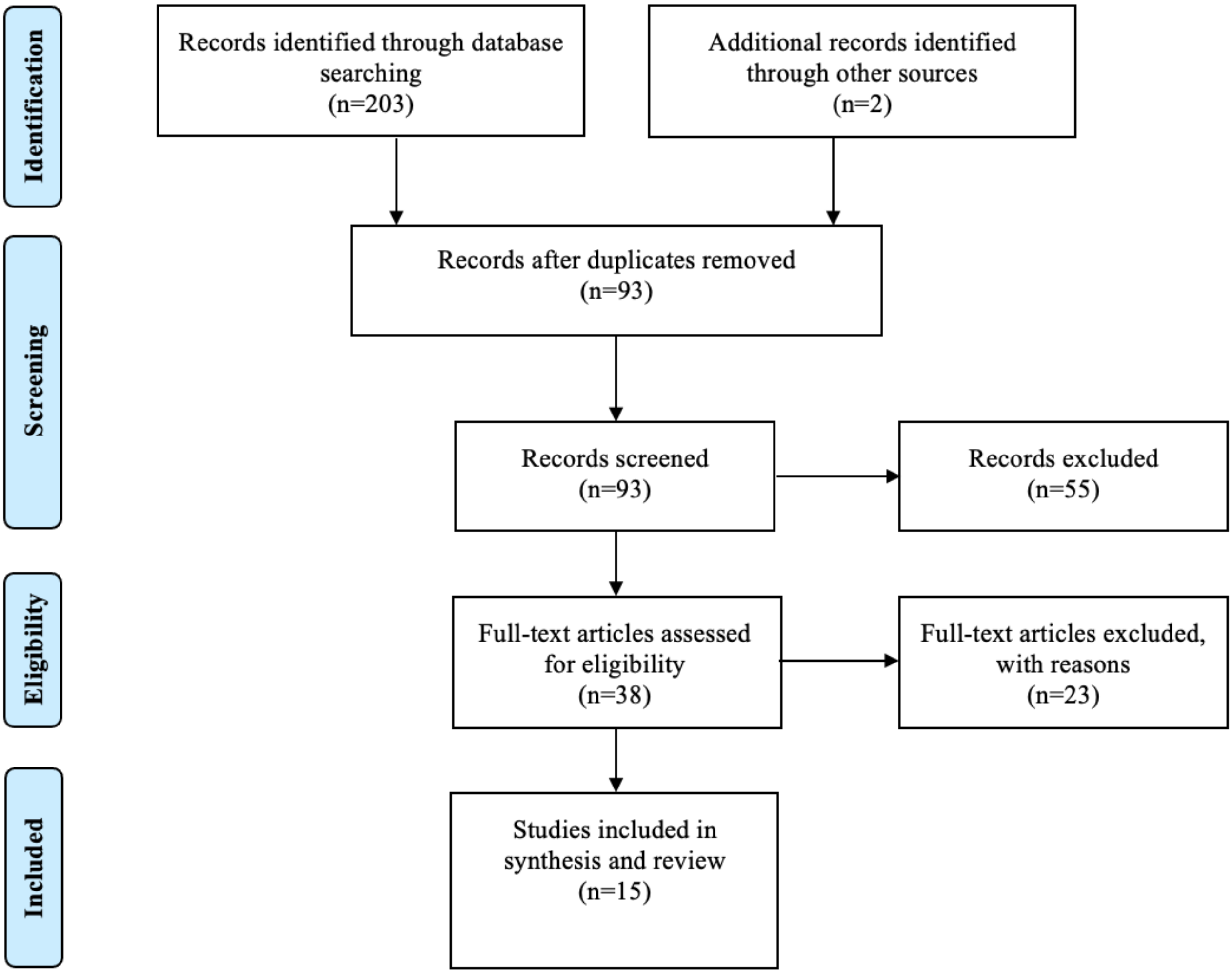
The literature assessing the effectiveness of telehealth practices remains ambiguous and many technology interventions are unverified. In rural communities, there are gaps in describing the design and effectiveness of technology interventions and best practices for preventing and treating specific diseases and addressing determinants of health. Synthesizing the current scope and application of telehealth is critical for guiding future interventions and policies. For this article, we defined telehealth as both telehealth and telemedicine and will include mobile health and digital health solutions: electronic health, telemedicine, artificial intelligence, electronic medical record/portal technology, videoconferencing, wearables and biosensors, and remote monitoring tools. The objective of this study was to review and evaluate literature published on the current applications, therapeutic areas, and outcomes of telehealth interventions in rural communities in the United States.

Methods

For this narrative review, we searched PubMed MEDLINE from January 2017 to December 2020. Key search terms included (“telehealth” AND “telemedicine”), “rural” AND “outcomes.” Filters were applied to identify free full-text studies published in English over a 4-year period to include the most recent evidence available in this research area. We included randomized controlled trials, mixed methods studies, qualitative studies, post hoc analyses, and prospective and retrospective cohort studies. Included studies contained at least one type of telehealth service and one primary measurable outcome and were assessed in rural communities/settings. We excluded systematic reviews, meta-analyses, editorials, conferences, unpublished studies and abstracts, studies outside the United States, and articles that were missing criteria based on the Critical Appraisal Skills Programme criteria. The Critical Appraisal Skills Programme has developed a set of 8 critical appraisal tools to assess the quality of evidence-based research; they have been widely used in previous studies [5].

To organize our review of the literature, we used Covidence, a review management tool, to conduct a review of titles and abstracts, and a full review of the articles. We used Covidence to exclude duplicate records. Both authors screened studies for relevance based on titles and abstracts. Both authors reviewed the full-text articles of relevant articles for study inclusion. Article discrepancies on study inclusion were resolved through formal discussion and consensus between the two authors. The Critical Appraisal Skills Programme checklists [6] were used to assess the quality and content of the articles. Additionally, we examined the reference lists of all included articles for other relevant references. Figure 1 outlines the article review process. We excluded articles due to irrelevance (n=13), wrong study design (n=8), wrong intervention (n=1), and wrong outcomes (n=1). We extracted the following information from each article: study design, telehealth type, therapeutic area, population, risk of bias, key message, and the primary outcome. We grouped the articles according to study population into the following categories: health care professionals, patients, and health care professionals and patients. This narrative literature review was exempt from ethics review as no human participant protection was required because no human participants were involved in this research.

Figure 1. PRISMA flow diagram. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.


Identification
Screening
Eligibility
Included
Records identified through database searching (n=203)
Additional records identified through other sources (n=2)
Records after duplicates removed (n=93)
Records screened (n=93)
Records excluded (n=55)
Full-text articles assessed for eligibility (n=38)
Full-text articles excluded, with reasons (n=23)
Studies included in synthesis and review (n=15)

Results

Overview

A total of 38 eligible articles were identified, of which 15 reported telehealth intervention outcomes and were included. The therapeutic areas examined included mental health (n=3), HIV (n=2), reproductive care/women's health (n=3), osteoporosis (n=1), orthopedics (n=1), acute ischemic stroke (n=1), cancer (n=1), substance use disorder (n=1), ophthalmology (n=1), and emergency medicine (n=1). The majority of the studies (n=9) analyzed telehealth interventions in patients, followed by health care professionals (n=3) and both patients and health care professionals (n=3). Of the patient-centered studies, two studies were specific to veterans, two to Medicare beneficiaries, and one to Medicaid beneficiaries. Table 1 outlines the characteristics the studies included in this review. The outcomes of these studies were focused on the following main themes: feasibility and acceptability of telehealth, diagnostic and treatment validation of telehealth, patient satisfaction and self-confidence, education and training, and telehealth design features including prevalence and access, type of service, and therapeutic area.

Table 1. Characteristics of the studies included in the review.

Study	Study design	Telehealth type	Therapeutic area	Outcomes	Risk of bias
Ward, 2018 [4]	Mixed methods	Tele-ED (emergency department)/telemedicine	Emergency medicine	Feasibility and acceptability	Medium
Hicken, 2017 [7]	Randomized	Internet and telephone-based care	Mental health	Feasibility and acceptability	Low
Stringer, 2018 [8]	Mixed methods	Electronic adherence monitor	HIV	Feasibility and acceptability	Low
Uscher-Pines, 2019 [9]	Randomized	Telelactation/telehealth	Reproductive care/women's health	Feasibility and acceptability	None
Huskamp, 2018 [10]	Retrospective	Tele–substance use disorder/telemedicine	Substance use disorder	Telehealth use	None
Mehrotra, 2017 [11]	Retrospective	Telemental/telemedicine	Mental health	Telehealth use	None
Sinha, 2019 [12]	Qualitative	Telemedicine	Orthopedics	Patient satisfaction	Low
Brecthel, 2018 [13]	Retrospective	Telestroke Network	Acute ischemic stroke	Diagnostic validation	Low
Kapinos, 2019 [14]	Post hoc analysis	Telelactation	Reproductive care/women's health	Design and demand	None
Talbot, 2019 [15]	Retrospective	Telehealth	Mental health	Prevalence, diagnosis, and type of service	Low
Lewiecki, 2017 [16]	Prospective	TeleECHO	Osteoporosis	Acceptability and self-confidence	Medium
Moeckli, 2017 [17]	Mixed methods	Extension for Community Health Outcomes (ECHO)/telemedicine	HIV	Application and acceptability	Low
Gilbertson-White, 2019 [18]	Mixed methods	Oncology Associated Symptoms and Individualized Strategies (OASIS)	Cancer	Patient needs and satisfaction	Low
Liu, 2019 [19]	Qualitative	Teleophthalmology	Ophthalmology	Feasibility and acceptability	None
Demirci, 2018 [20]	Qualitative	Telelactation/telemedicine	Reproductive care/women's health	Feasibility and acceptability	Low

Patients

Of the nine studies that analyzed telehealth interventions in patients, three (33%) of them examined the feasibility and acceptability of telehealth. These studies demonstrated that internet, electronic adherence monitors, and telelactation are feasible and accepted in rural, underserved populations and improve access to care (Table 2). The disadvantages were the following: comparative effectiveness outcomes were not different between caregivers receiving technology interventions and those receiving telephone-delivered support, results were not statistically significant in detecting differences in breastfeeding duration and exclusivity, and technological difficulties, such as loss of connectivity [7-9].

Table 2. Key messages from included studies.

Study	Key message
Ward, 2018 [4]	Telemedicine led to decreased staffing costs and improved physician recruitment and retention.
Hicken, 2017 [7]	Technology demonstrated feasibility and acceptability for delivering caregiver support to a group of largely older, rural, spousal caregivers of veterans with dementia.
Stringer, 2018 [8]	Electronic adherence monitor is acceptable and feasible in a rural US setting, but technological difficulties were present and may impede effectiveness.
Uscher-Pines, 2019 [9]	Telelactation participants were breastfeeding at higher rates and telelactation can be implemented in a rural underserved population.
Huskamp, 2018 [10]	There were low use rates of tele–substance use disorder (Tele-SUD) overall. Future studies should evaluate the effect of Tele-SUD on access and outcomes.
Mehrotra, 2017 [11]	States with a telemedicine law and a pro–telemental health regulatory environment had significantly higher rates of telemental health use.
Sinha, 2019 [12]	Telemedicine visits decreased indirect and direct costs, reduced travel time, and resulted in similar patient satisfaction.
Brecthel, 2018 [13]	Telestroke provides less restrictive criteria for clinical risk factors associated with the inclusion of hypertensive patients with stroke for thrombolysis.
Kapinos, 2019 [14]	Telelactation showed both demand for and positive experiences with telelactation in an underserved population.
Talbot, 2019 [15]	Rural Medicaid beneficiaries were more likely to use telehealth services than their urban counterparts, but absolute rates of telehealth use were low.
Lewiecki, 2017 [16]	TeleECHO showed substantial improvement of self-confidence in 20 domains of osteoporosis care and can improve osteoporosis care with greater convenience and lower cost than referral to a specialty center.
Moeckli, 2017 [17]	There was limited uptake of HIV Extension for Community Health Outcomes (ECHO) telemedicine in settings where veterans traveled to distant specialty clinics. Other telemedicine models should be considered for HIV care.
Gilbertson-White, 2019 [18]	Oncology Associated Symptoms and Individualized Strategies (OASIS) is easy to use, contains relevant content, and has pleasing graphics. Rural stakeholders perceived OASIS positively.
Liu, 2019 [19]	Patients and primary care providers have limited familiarity with teleophthalmology for diabetic eye screening and primary care providers reported difficulties with use.
Demirci, 2018 [20]	Telelactation was convenient and efficient, was accepted in rural areas lacking breastfeeding support services, increased maternal breastfeeding confidence, and showed several advantages over in-person and telephone-based support. Telelactation appears to be an acceptable delivery model for lactation assistance in rural areas.

A total of two studies examined the outcomes of telehealth use in Medicare patients with substance use disorder (SUD) and mental health disorders, designated as Tele-SUD and telemental, respectively. Huskamp et al [10] concluded that Tele-SUD has low use rates and is primarily used to complement in-person care and is disproportionately used by those with relatively severe SUD. Mehrotra et al [11] concluded that beneficiaries who received a telemental visit were more likely to be younger than 65 years old, be eligible for Medicare because of disability, and live in a relatively poor community; in addition, states with a pro–telemental health regulatory environment had significantly higher rates of telemental health use than those that did not.

The outcomes of the other four patient-centered studies included patient satisfaction, diagnostic validation, design and demand, and prevalence, diagnosis, and type of service. The advantages of telemedicine visits found in these studies included the following: decreased indirect and direct costs, lower travel costs and travel times, similar patient satisfaction compared to onsite visits, and patient satisfaction among telelactation users [12-14]. In addition, telestroke technology provides less restrictive criteria for clinical risk factors associated with the inclusion of hypertensive patients with stroke for thrombolysis [13]. Talbot et al [15] concluded that rural Medicaid beneficiaries were more likely to use telehealth services than their urban counterparts, psychotropic medication management was the most prevalent use of telehealth, the proportion of users who accessed nonbehavioral health services through telehealth was significantly greater as rurality increased, and significantly higher proportions of telehealth users received services to address attention deficit hyperactivity disorder. There were no direct disadvantages reported in these four studies.

Health Care Professionals

A total of three studies analyzed telehealth interventions in health care professionals and reported outcomes. Lewiecki et al [16] examined the acceptability of TeleECHO and found that self-confidence in 20 domains of osteoporosis care showed substantial improvement. In addition, they determined that TeleECHO can contribute to alleviating the osteoporosis care crisis by leveraging scarce resources; providing motivated practitioners with the skills to provide better skeletal health care, closer to home, with greater convenience; and being lower cost than referral to a specialty center. Additionally, TeleECHO can be applied to any location worldwide with internet access, allowing access in local time zones and a variety of languages [16].

Another study examined the application and acceptability of a telemedicine intervention called HIV ECHO. Moeckli et al [17] showed a limited adoption of ECHO, which was attributed partly to shifting ownership of care from HIV specialists to primary care providers (PCPs) and low HIV prevalence and long treatment cycles that prevented rapid learning loops for PCPs. More specifically, there was limited uptake of HIV ECHO telemedicine programs in settings where veterans historically traveled to distant specialty clinics [17]. The third study evaluated the feasibility and acceptability of a technology intervention for emergency departments, Tele-ED. Ward et al [4] concluded that Tele-ED hospitals tended to have decreased ED staffing costs—while the hospitals not applying this policy showed continually increasing staffing costs over time—and improved physician recruitment and retention. The only disadvantage to the study was limited uptake of Tele-ED (7/19 hospitals, 37%); however, these results conclude that more hospitals will likely use telemedicine to provide physician backup for advanced practice providers staffing the ED [4].

Patients and Health Care Professionals

A total of three studies analyzed telehealth interventions in patients and health care professionals. Gilbertson-White et al [18] examined patient needs and satisfaction with an electronic health tool, Oncology Associated Symptoms and Individualized Strategies (OASIS), and concluded that the web application is easy to use, contains relevant content, has pleasing graphics, and was perceived positively. There were infrequent users of OASIS in the group; however, both frequent and infrequent internet users positively evaluated the web application [18]. Liu et al [19] tested the feasibility and acceptability of teleophthalmology and concluded that patients and PCPs have limited familiarity with teleophthalmology for diabetic eye screening. A major disadvantage to teleophthalmology was that PCPs reported significant difficulty identifying when patients are due for diabetic eye screening and could not sufficiently initiate referrals [19].

Another study examined the feasibility and acceptability of telelactation. The advantages of telelactation were as follows: convenient and efficient, provided a needed service in rural areas lacking breastfeeding support services, and increased maternal breastfeeding confidence [20]. The barriers to use included maternal reluctance to conduct video calls with an unknown provider, preference for community-based breastfeeding resources, and technical issues, including limited Wi-Fi in rural areas [20].

Discussion

Principal Findings

This review of the literature indicates telehealth is used for a variety of disease states and rural populations across the United States. Health technology interventions are a critical component of health care services in rural areas; they decrease staffing costs, travel costs, and travel time, and increase the ability of residents to seek care (including specialty care) that they otherwise would not be able to access in remote locations. This review explored telehealth in a broad sense and included technology models for clinical use, education and training of health care professionals and patients, and preventive and primary care services. All of these included models have shown feasibility and acceptability in rural populations, validating their importance and potential to improve outcomes and access to care.

Overall, the included studies reported positive outcomes and experiences of telehealth use in rural populations, including acceptability and increased satisfaction; in addition, the technology was considered convenient and efficient. Other notable benefits included decreased direct and indirect costs to the patient (travel cost and time) and health care service provider (staffing), lower onsite health care resource utilization, improved physician recruitment and retention, improved access to care, and increased education and training of patients and health care professionals.

Disadvantages of telehealth interventions included having tele-visits with unknown providers and technological issues such as loss of connectivity and limited Wi-Fi access in rural areas. Several studies reported that comparative effectiveness outcomes between telehealth and traditional visits were not statistically significant; however, these studies also noted telehealth technology was well accepted and implemented in rural, underserved populations and described the importance of testing additional technology interventions in these populations to identify which telehealth programs are most effective.

Feasibility and acceptability are the foundation for implementing new health technologies in any environment or setting. This validation shows that rural populations and underserved communities have the capability to implement telehealth, report satisfaction with telehealth interventions, and describe the interventions as both convenient and effective. Even in articles where telehealth programs did not show statistical significance, these programs have demonstrated feasibility and acceptability. Limitations often associated with technologies (eg, lack of Wi-Fi in remote locations, connectivity issues, and inability of persons to use technology successfully) were reported at low rates, much lower than common perceptions about using these technologies in rural communities. This shows these technologies can be implemented successfully and supports the extension of telehealth programs in rural communities throughout the United States.

A key additional finding of this review regards the appropriate types of telehealth programs. Although these technologies have been verified, these studies consistently highlight the need to test various telehealth programs in specific communities to validate which models are most effective; for example, diagnostic telemedicine versus telehealth video calls will have greater benefits in different settings and populations. It is important that we go beyond the scope of feasibility and acceptability to make sure we use telehealth programs to their full benefit and effect. The benefits have been established but the possibility of expansion of telehealth programs does not rely solely on testing just one model for a population but finding which model fits the community the best.

Additionally, while no articles with a primary outcome of telehealth use for COVID-19 met the criteria for inclusion in this review, the landscape of telehealth and digital health care has changed dramatically since the start of the pandemic. The acceptance of telehealth during the pandemic has accelerated to the point that many of the barriers to telehealth use may have disappeared. To our knowledge, no reviews have been published on this topic in PubMed MEDLINE. Future reviews should focus on the changes in and outcomes of telehealth use in rural communities across the United States due to the COVID-19 pandemic.

The administrative regulations behind telehealth programs, especially in areas where reimbursement for telehealth programs is not enforced by the state, should be highly considered. Requiring insurance companies to provide telehealth services to their patients is key to providing access to health care services in rural communities. Rural communities lack the resources, staff, and expertise to be able to positively affect health care outcomes, health care quality, and health equity. This review highlights the need for additional telehealth program studies and research into long-term real-world outcomes. Evidence-based studies have the potential to establish significance and comparative effectiveness against traditional health care (ie, onsite services). Additionally, state, federal, and local policy should be updated to cover the use of these programs and provide grants and funding for researchers to implement and test these programs in rural, underserved communities to improve access and quality of care.

Limitations

There are several limitations to this narrative review. Publication bias is possible within this study as we leveraged only PubMed MEDLINE and omitted grey literature such as reports, government documents and releases, working papers, white papers, and evaluations. Searching additional databases would potentially provide additional articles to be in included in this review. Additionally, limiting our search time frame to the past four years excludes earlier publications and data on health technology interventions and outcomes; however, the objective of this review was to report the most recent information on telehealth programs as they have advanced and expanded greatly over the past several years. Lastly, this review only included rural communities in the United States and would not be generalizable to non-US territories or domestic, urban communities, and populations affected by COVID-19. This study is generalizable to rural, underserved populations and potentially to the clinical settings and specific therapeutic areas studied in the included articles.

Conclusion

This review highlights the current scope of using telehealth interventions in rural populations across the United States. Telehealth models were associated with positive outcomes for patients and health care professionals, suggesting these models can be effective for continuing education and training in the workplace. The findings of this review are limited to rural, domestic communities and are concentrated in specific therapeutic areas of disease. The findings support the existing literature on the need to increase and validate telehealth interventions and further update and implement policies to increase access and provide high-quality telehealth programs. Future telehealth interventions and studies examining these programs are warranted, especially in rural communities, and future research should evaluate the impact of increased telehealth use as a result of the COVID-19 pandemic.

Authors' Contributions

MB was the principal study investigator and conducted the article search. MB and YC reviewed all articles for inclusion. MB wrote the first draft of the article with input from YC. Both authors were involved in the study design, contributed to the interpretation of the results, contributed to the writing and revision of the manuscript, and approved the decision to submit the article for publication.

Conflicts of Interest

None declared.

References

1. Orlando JF, Beard M, Kumar S. Systematic review of patient and caregivers' satisfaction with telehealth videoconferencing as a mode of service delivery in managing patients' health. PLoS One 2019;14(8):e0221848 [FREE Full text] [doi: 10.1371/journal.pone.0221848] [Medline: 31469865]
2. ter Huurne ED, Postel MG, de Haan HA, Drossaert CHC, DeJong CAJ. Web-based treatment program using intensive therapeutic contact for patients with eating disorders: before-after study. J Med Internet Res 2013 Feb 04;15(2):e12 [FREE Full text] [doi: 10.2196/jmir.2211] [Medline: 23380291]
3. Kruse CS, Krowski N, Rodriguez B, Tran L, Vela J, Brooks M. Telehealth and patient satisfaction: a systematic review and narrative analysis. BMJ Open 2017 Aug 03;7(8):e016242 [FREE Full text] [doi: 10.1136/bmjopen-2017-016242] [Medline: 28775188]
4. Ward MM, Merchant KAS, Carter KD, Zhu X, Ullrich F, Wittrock A, et al. Use Of Telemedicine For ED Physician Coverage In Critical Access Hospitals Increased After CMS Policy Clarification. Health Aff (Millwood) 2018 Dec;37(12):2037-2044. [doi: 10.1377/hlthaff.2018.05103] [Medline: 30633684]
5. CASP Checklist. Critical Appraisal Skills Programme. URL: https://casp-uk.b-cdn.net/wp-content/uploads/2018/01/CASP-Bibliography.pdf [accessed 2018-01-09]
6. CASP Checklist. Critical Appraisal Skills Programme. URL: https://casp-uk.net/casp-tools-checklists/ [accessed 2018-01-09]

7. Hicken BL, Daniel C, Luptak M, Grant M, Kilian S, Rupper RW. Supporting Caregivers of Rural Veterans Electronically (SCORE). J Rural Health 2017 Jun;33(3):305-313. [doi: 10.1111/jrh.12195] [Medline: 27437642]
8. Stringer KL, Azuero A, Ott C, Psaros C, Jagielski CH, Safren SA, et al. Feasibility and Acceptability of Real-Time Antiretroviral Adherence Monitoring among Depressed Women Living with HIV in the Deep South of the US. AIDS Behav 2019 May;23(5):1306-1314 [FREE Full text] [doi: 10.1007/s10461-018-2322-z] [Medline: 30377982]
9. Uscher-Pines L, Ghosh-Dastidar B, Bogen DL, Ray KN, Demirci JR, Mehrotra A, et al. Feasibility and Effectiveness of Telelactation Among Rural Breastfeeding Women. Acad Pediatr 2020 Jul;20(5):652-659. [doi: 10.1016/j.acap.2019.10.008] [Medline: 31629118]
10. Huskamp HA, Busch AB, Souza J, Uscher-Pines L, Rose S, Wilcock A, et al. How Is Telemedicine Being Used In Opioid And Other Substance Use Disorder Treatment? Health Aff (Millwood) 2018 Dec;37(12):1940-1947 [FREE Full text] [doi: 10.1377/hlthaff.2018.05134] [Medline: 30633671]
11. Mehrotra A, Huskamp HA, Souza J, Uscher-Pines L, Rose S, Landon BE, et al. Rapid Growth In Mental Health Telemedicine Use Among Rural Medicare Beneficiaries, Wide Variation Across States. Health Aff (Millwood) 2017 May 01;36(5):909-917. [doi: 10.1377/hlthaff.2016.1461] [Medline: 28461359]
12. Sinha N, Cornell M, Wheatley B, Munley N, Seeley M. Looking Through a Different Lens: Patient Satisfaction With Telemedicine in Delivering Pediatric Fracture Care. J Am Acad Orthop Surg Glob Res Rev 2019 Sep;3(9):e100 [FREE Full text] [doi: 10.5435/JAAOSGlobal-D-19-00100] [Medline: 31773080]
13. Brecthel L, Gainey J, Penwell A, Nathaniel TI. Predictors of thrombolysis in the telestroke and non telestroke settings for hypertensive acute ischemic stroke patients. BMC Neurol 2018 Dec 21;18(1):215 [FREE Full text] [doi: 10.1186/s12883-018-1204-3] [Medline: 30577762]
14. Kapinos K, Kotzias V, Bogen D, Ray K, Demirci J, Rigas MA, et al. The Use of and Experiences With Telelactation Among Rural Breastfeeding Mothers: Secondary Analysis of a Randomized Controlled Trial. J Med Internet Res 2019 Sep 03;21(9):e13967 [FREE Full text] [doi: 10.2196/13967] [Medline: 31482848]
15. Talbot JA, Burgess AR, Thayer D, Parenteau L, Paluso N, Coburn AF. Patterns of Telehealth Use Among Rural Medicaid Beneficiaries. J Rural Health 2019 Jun;35(3):298-307. [doi: 10.1111/jrh.12324] [Medline: 30288808]
16. Lewiecki EM, Rochelle R, Bouchonville MF, Chafey DH, Olenginski TP, Arora S. Leveraging Scarce Resources With Bone Health TeleECHO to Improve the Care of Osteoporosis. J Endocr Soc 2017 Dec 01;1(12):1428-1434 [FREE Full text] [doi: 10.1210/js.2017-00361] [Medline: 29264466]
17. Moeckli J, Stewart KR, Ono S, Alexander B, Goss T, Maier M, et al. Mixed-Methods Study of Uptake of the Extension for Community Health Outcomes (ECHO) Telemedicine Model for Rural Veterans With HIV. J Rural Health 2017 Jun;33(3):323-331. [doi: 10.1111/jrh.12200] [Medline: 27557039]
18. Gilbertson-White S, Yeung CW, Saeidzadeh S, Tykol H, Vikas P, Cannon A. Engaging Stakeholders in the Development of an eHealth Intervention for Cancer Symptom Management for Rural Residents. J Rural Health 2019 Mar;35(2):189-198. [doi: 10.1111/jrh.12297] [Medline: 29696696]
19. Liu Y, Zupan NJ, Swearingen R, Jacobson N, Carlson JN, Mahoney JE, et al. Identification of barriers, facilitators and system-based implementation strategies to increase teleophthalmology use for diabetic eye screening in a rural US primary care clinic: a qualitative study. BMJ Open 2019 Feb 18;9(2):e022594 [FREE Full text] [doi: 10.1136/bmjopen-2018-022594] [Medline: 30782868]
20. Demirci J, Kotzias V, Bogen DL, Ray KN, Uscher-Pines L. Telelactation via Mobile App: Perspectives of Rural Mothers, Their Care Providers, and Lactation Consultants. Telemed J E Health 2019 Sep;25(9):853-858 [FREE Full text] [doi: 10.1089/tmj.2018.0113] [Medline: 30212280]

Abbreviations

ECHO: Extension for Community Health Outcomes
ED: emergency department
OASIS: Oncology Associated Symptoms and Individualized Strategies
PCP: primary care provider
SUD: substance use disorder

Edited by R Kukafka; submitted 12.04.21; peer-reviewed by C Hawke, G Margelis; comments to author 19.05.21; revised version received 07.06.21; accepted 05.07.21; published 26.08.21.

Please cite as:
Butzner M, Cuffee Y
Telehealth Interventions and Outcomes Across Rural Communities in the United States: Narrative Review
J Med Internet Res 2021;23(8):e29575
URL: https://www.jmir.org/2021/8/e29575
doi:10.2196/29575
PMID:34435965

©Michael Butzner, Yendelela Cuffee. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Using a Smartphone Application for the Accurate and Rapid Diagnosis of Acute Anterior Intracranial Arterial Occlusion: Usability Study

Teppei Komatsu[1], MD, PhD; Kenichiro Sakai[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:
Teppei Komatsu, MD, PhD
Department of Neurology
The Jikei University School of Medicine
3-25-8 Nishishimbashi
Minato-ku
Tokyo, 105-8461
Japan
Phone: 81 334331111
Fax: 81 335788071
Email: teppeinoieni.coicoi@jikei.ac.jp

Abstract

Background: Telestroke has developed rapidly as an assessment tool for patients eligible for reperfusion therapy.

Objective: To investigate whether vascular neurologists can diagnose intracranial large vessel occlusion (LVO) as quickly and accurately using a smartphone application compared to a hospital-based desktop PC monitor.

Methods: We retrospectively enrolled 108 consecutive patients with acute ischemic stroke in the middle cerebral artery territory who underwent magnetic resonance imaging (MRI) within 24 hours of their stroke onset. Two vascular neurologists, blinded to all clinical information, independently evaluated magnetic resonance angiography and fluid-attenuated inversion recovery images for the presence or absence of LVO in the internal carotid artery and middle cerebral artery (M1, M2, or M3) on both a smartphone application (Smartphone-LVO) and a hospital-based desktop PC monitor (PC-LVO). To evaluate the accuracy of an arterial occlusion diagnosis, interdevice variability between Smartphone-LVO and PC-LVO was analyzed using κ statistics, and image interpretation time was compared between Smartphone-LVO and PC-LVO.

Results: There was broad agreement between Smartphone-LVO and PC-LVO evaluations regarding the presence or absence of arterial occlusion (Reader 1: κ=0.94; *P*<.001 vs Reader 2: κ=0.89; *P*<.001), and interpretation times were similar between Smartphone-LVO and PC-LVO.

Conclusions: The results indicate the evaluation of neuroimages using a smartphone application can provide an accurate and timely diagnosis of anterior intracranial arterial occlusion that can be shared immediately with members of the stroke team to support the management of patients with hyperacute ischemic stroke.

(J Med Internet Res 2021;23(8):e28192) doi:10.2196/28192

KEYWORDS

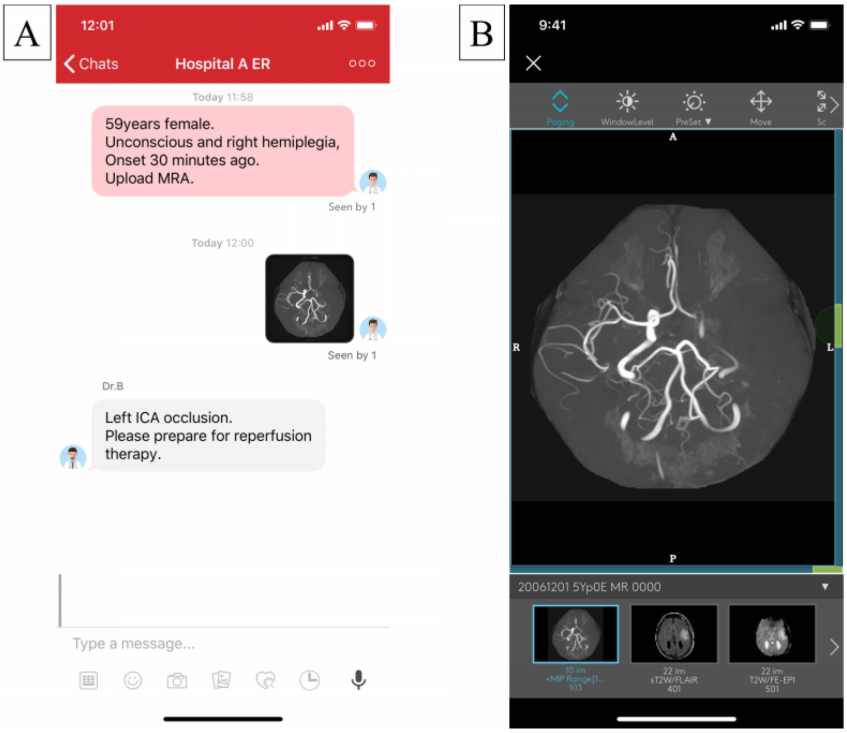
stroke; infarction; teleradiology; smartphone; telehealth; reperfusion; neurology; mHealth; application; mobile health; mobile applications; diagnosis; diagnostics

Introduction

In acute stroke management, early initiation is the most important factor in successful reperfusion therapy for both intravenous tissue plasminogen activator (IV-tPA) and mechanical thrombectomy (MT) [1,2] because the patient loses 1.9 million neurons every minute that a stroke is untreated [3]. The selection of appropriate candidates for reperfusion therapy requires neurological evaluation and neuroimaging studies to be obtained immediately after admission. Accurate diagnostic information must be shared promptly within the multidisciplinary stroke team (comprising of emergency physicians, vascular neurologists, neuroradiologists, neurosurgeons, anesthetists, and ward nurses) to progress the decision-making process.

Because “time is brain” in acute stroke treatment, the appropriate use of information and communication technology (ICT), particularly telemedicine for stroke or “telestroke,” has become a valuable method for reaching out to patients who are eligible for IV-tPA and MT administration. Telestroke is continually undergoing development, specifically in the United States and the European Union. The JOIN smartphone application enables the multidisciplinary stroke team to share clinical information and imaging data securely. Through JOIN, the intrahospital network system sends medical images from the hospital server to a shared chat room accessible to stroke team members only, similar to modified teleradiology and teleconferencing (Figure 1). Over 200 medical institutions in Japan and more than 11 other countries, including the United States, Germany, and Brazil, use JOIN. It is complementary to the telestroke system, particularly as an image evaluation tool [4]. In addition, the COVID-19 epidemic has increased the importance of teleradiology.

Figure 1. A. The JOIN smartphone application utilizes the easy-to-use interface of the social networking communication environment. B. The JOIN smartphone application displays diagnostic medical images, such as magnetic resonance imaging. ICA: internal carotid artery; MRA: magnetic resonance angiography.



59years female.
Unconscious and right hemiplegia,
Onset 30 minutes ago.
Upload MRA.

Left ICA occlusion.
Please prepare for reperfusion therapy.

Because the evaluation of large vessel occlusions (LVO) is vital for successful MT [5-7], the accurate and rapid diagnosis of LVO is clinically relevant. ICT can play an essential role in this regard by providing intranetwork and internetwork support. At our stroke center, the patients with acute ischemic stroke are initially examined via magnetic resonance imaging (MRI) rather than computed tomography (CT) due to hyperacute ischemic stroke being easily diagnosed on an MRI. LVO can be assessed via magnetic resonance angiography (MRA) without the use of a contrast agent. Although a previous study has reported favorable interdevice agreement between smartphone applications and desktop PC monitoring systems for evaluating CT images [8] and diffusion-weighted imaging on MRI [9], interdevice agreement for LVO on MRI is unknown. This study compares the speed and accuracy of LVO diagnosis by vascular neurologists between the JOIN smartphone application (Smartphone-LVO) and a conventional hospital-based image viewer using a desktop PC monitor (PC-LVO).

Methods

Patients

Consecutive patients admitted with ischemic stroke to Jikei University Hospital, Japan, between January 2016 and September 2017 were retrospectively enrolled in the study. The subtype of ischemic stroke was classified according to the TOAST (Trial of ORG 10172 in Acute Stroke Treatment) [10]. Stroke severity on admission was assessed using the National Institutes of Health Stroke Scale score [11]. After reviewing patients’ medical records, 108 patients with ischemic stroke in the middle cerebral artery territory identified on MRI within 24 hours of onset were enrolled in the study. Patients who did not undergo MRI upon admission or for whom imaging was not available were excluded from the study.

JOIN Smartphone Application

The JOIN (Allm, Tokyo, Japan) smartphone application enables teleradiology of medical images, including MRI, CT, ultrasonography, and angiography (Figure 1). Using JOIN, medical information can be shared immediately among the stroke team via SMS. Images from the initial MRI brain examination are sent immediately from the picture archiving and communication system (PACS) on the hospital server to each smartphone on which JOIN is installed. Users can enlarge and evaluate the images through a simple touch sequence on the screen of their smartphone.

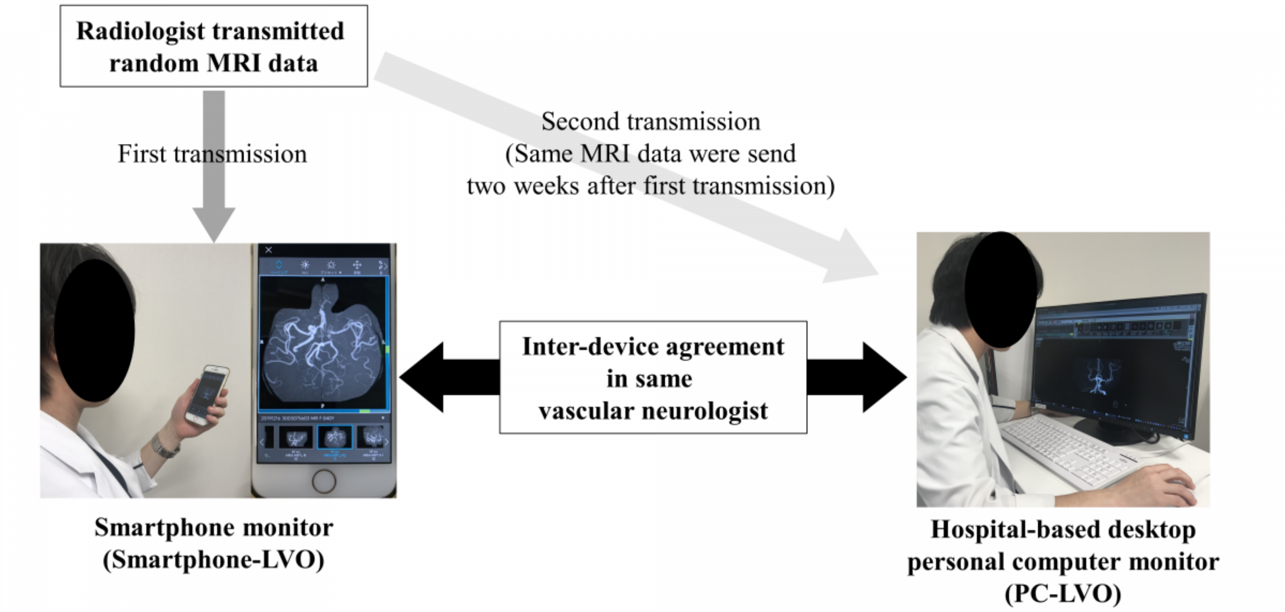
MRI Protocols

All patients underwent routine 3D time-of-flight (TOF) MRA of the intracranial arteries and fluid-attenuated inversion recovery (FLAIR) of the brain on a 1.5T scanner (MAGNETOM Avanto and MAGNETOM Symphony). The imaging parameters were TOF MRA: repetition time (TR)/ echo time (TE) = 23/7.15 ms, flip angle = 20°, section thickness = 0.6 mm, matrix = 320 × 288, field-of-view (FOV) = 18 cm; FLAIR: TR/TE = 9000/94 ms, flip angle = 160°, section thickness = 5 mm, section gap = 1.5 mm, matrix = 256 × 256, FOV = 21 cm.

Image Review

The image review process is shown in Figure 2. Prior to the review, 2 readers (TK and KS, Readers 1 and 2 with 8 years and 10 years of experience in vascular neurology, respectively) installed JOIN on a Jikei University Hospital iPhone (version 7; Apple). After they confirmed the operation of the application, a radiologist transmitted the MRA and FLAIR data to the application. Patient image sets were presented randomly, and the readers were blinded to all background and clinical information. Both readers evaluated all the images from all patients. After receiving the MRA and FLAIR images, the readers independently assessed them on the JOIN application for the presence or absence of intracranial arterial occlusion; in the case of LVO, the site of arterial occlusion was identified. In patients with acute stroke, it is important to check for hyperintense vessel sign (HVS) on FLAIR, which suggests the presence of LVO [12]. HVS occurs because of slow flow due to LVO but is not an indicator of infarction. Therefore, the 2 readers evaluated the MRA images for LVO and evaluated the FLAIR images for HVS. The arterial occlusion site was classified as internal carotid artery (ICA), the horizontal segment of the middle cerebral artery (M1), or insular or cortical segments of the middle cerebral artery (M2 or M3, respectively). The diagnosis accuracy of arterial occlusion was defined as interdevice variability between Smartphone-LVO and PC-LVO. We recorded the interpretation time (minutes), defined as the time between receiving the MRI data to completing the interpretation. We used 1 to 4 min to interpret times in segments of 0 to 1 min, 1 to 2 min, 2 to 3 min, and 3 to 4 min. The 2 readers evaluated all MRI data using the hospital desktop PC monitor in the same manner 2 weeks later.

Figure 2. Flow diagram of the study design. LVO: large vessel occlusion, MRI: magnetic resonance imaging.


Radiologist transmitted random MRI data
First transmission
Second transmission (Same MRI data were send two weeks after first transmission)
Inter-device agreement in same vascular neurologist
Smartphone monitor (Smartphone-LVO)
Hospital-based desktop personal computer monitor (PC-LVO)

Statistical Analysis

Interdevice agreement for evaluating the presence of arterial occlusion was analyzed between Smartphone-LVO and PC-LVO via kappa statistics. Kappa scores for agreement were rated as follows: poor (<0.20), fair (0.21-0.40), moderate (0.41-0.60), favorable (0.61-0.80), almost perfect (0.81-1.0).

Wilcoxon signed-rank test was used to compare image interpretation times between Smartphone-LVO and PC-LVO. *P* values <.05 indicated statistical significance. All analyses were performed using SPSS for Windows (version 22.0; BM-Armonk).

Standard Protocol Approvals and Registrations

The Regional Ethics and Hospital Management Committee of the Jikei University School of Medicine approved the study (approval No. 29-197). The board waived the need for patient consent by giving patients from whom data had been collected the opportunity to opt-out of the study.

Results

Patient Characteristics

Table 1 lists the patient characteristics. We enrolled 108 patients with a median age of 69 years and a median NIHSS score of 4 upon admission, of whom 72 (66%) were male.

Table 1. Patient characteristics.

Variable	All patients (n=108)
Age (years), median (IQR)	69 (61-80)
Sex (male), n (%)	72 (67)
NIHSS[a] score on admission, median (IQR)	4 (2-7)
TOAST[b] classification, n (%)	
Large-artery atherosclerosis	8 (7)
Small-vessel occlusion	14 (13)
Cardio-embolism	35 (32)
Other determined etiology	9 (8)
Undetermined	42 (39)
Modified Rankin scale at 3 months, median (IQR)	1 (1-3)
MRI[c] time from onset (minutes), median (IQR)	244 (107-623)

[a] NIHSS: National Institute Health Stroke Scale.

[b] TOAST: Trial of ORG 10172 in Acute Stroke Treatment.

[c] MRI: magnetic resonance imaging.

Interdevice Agreement Between Smartphone-LVO and PC-LVO

Interdevice agreement for evaluating the presence or absence of intracranial arterial occlusion was almost perfect between Smartphone-LVO and PC-LVO (Reader 1: κ=0.94, 95% CI 0.82-0.97; *P*<.001 and Reader 2: κ=0.89, 95% CI 0.77-0.92; *P*<.001; Table 2; Figure 3). Regarding detection of lesions in individual arteries, interdevice agreement was almost perfect for ICA and M1 occlusion (Tables 3 and 4; Figure 3). For analyzing M2 or M3 occlusion, interdevice agreement by Reader 1 was almost perfect (κ=0.89, 95% CI 0.77-0.93; *P*<.001,) and that by Reader 2 was favorable (κ=0.75, 95% CI 0.60-0.80 *P*<.001; Table 5; Figure 3).

Table 2. Comparison of evaluation of the presence or absence of intracranial arterial occlusion between Smartphone-LVO and PC-LVO.

Smartphone-LVO[a]	PC-LVO	
	Occlusion	No occlusion
Reader 1 (TK)		
Occlusion	33	2
No occlusion	1	72
Reader 2 (KS)		
Occlusion	42	5
No occlusion	1	60

[a] LVO: large vessel occlusion.

Figure 3. Interdevice agreement for evaluation of intracerebral arterial occlusion and for the 3 vessel categories. The degree of interdevice agreement was almost perfect in the evaluation of the arterial occlusion sites and was higher for proximal than distal arterial occlusion sites. ICA: internal carotid artery.

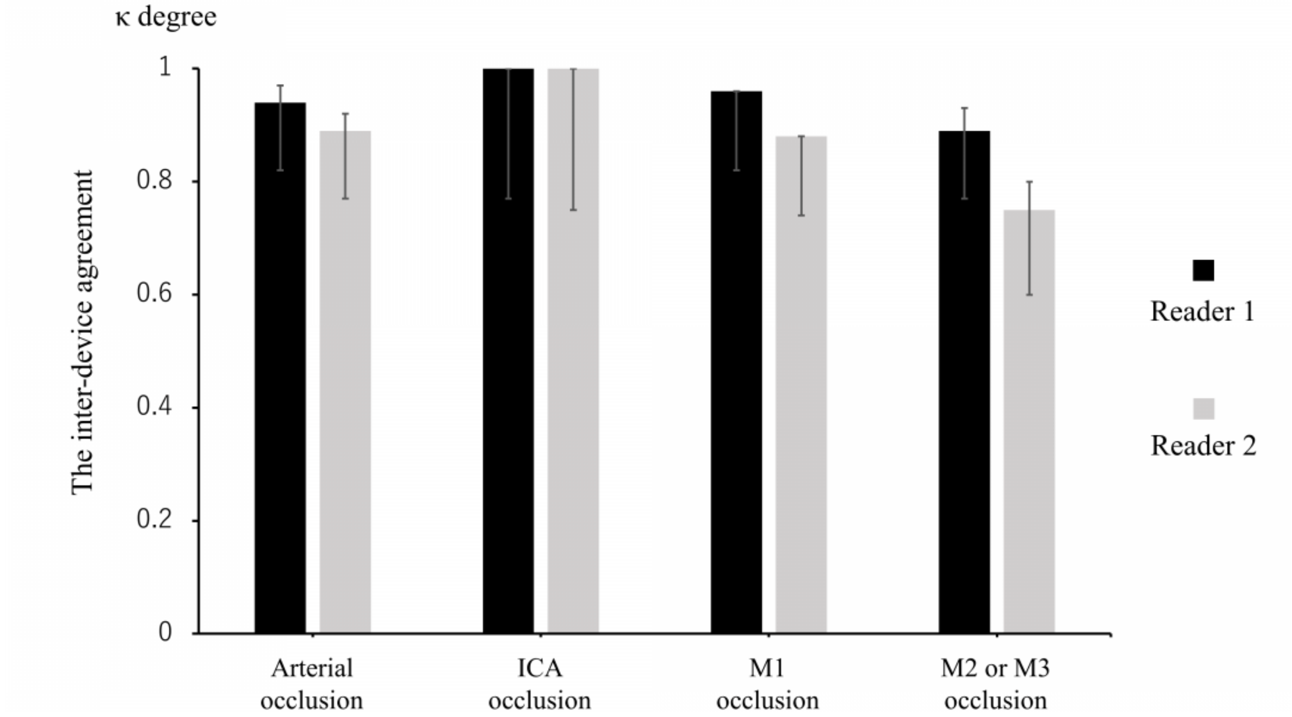


Table 3. Comparison of identification of internal carotid artery occlusion between Smartphone-LVO and PC-LVO.

Smartphone-LVO[a]	PC-LVO	
	Occlusion	No occlusion
Reader 1 (TK)		
Occlusion	8	0
No occlusion	0	100
Reader 2 (KS)		
Occlusion	7	0
No occlusion	0	101

[a] LVO: large vessel occlusion.

Table 4. Comparison of identification of horizontal segment of middle cerebral artery occlusion between Smartphone-LVO and PC-LVO.

Smartphone-LVO[a]	PC-LVO	
	Occlusion	No occlusion
Reader 1 (TK)		
Occlusion	16	1
No occlusion	0	91
Reader 2 (KS)		
Occlusion	19	0
No occlusion	4	85

[a] LVO: large vessel occlusion.

Table 5. Comparison of identification of insular or cortical segments of middle cerebral artery occlusion between Smartphone-LVO and PC-LVO.

Smartphone-LVO[a]	PC-LVO	
	Occlusion	No occlusion
Reader 1 (TK)		
Occlusion	31	4
No occlusion	1	72
Reader 2 (KS)		
Occlusion	34	11
No occlusion	2	61

[a] LVO: large vessel occlusion.

Interpretation Time Between Smartphone-LVO and PC-LVO

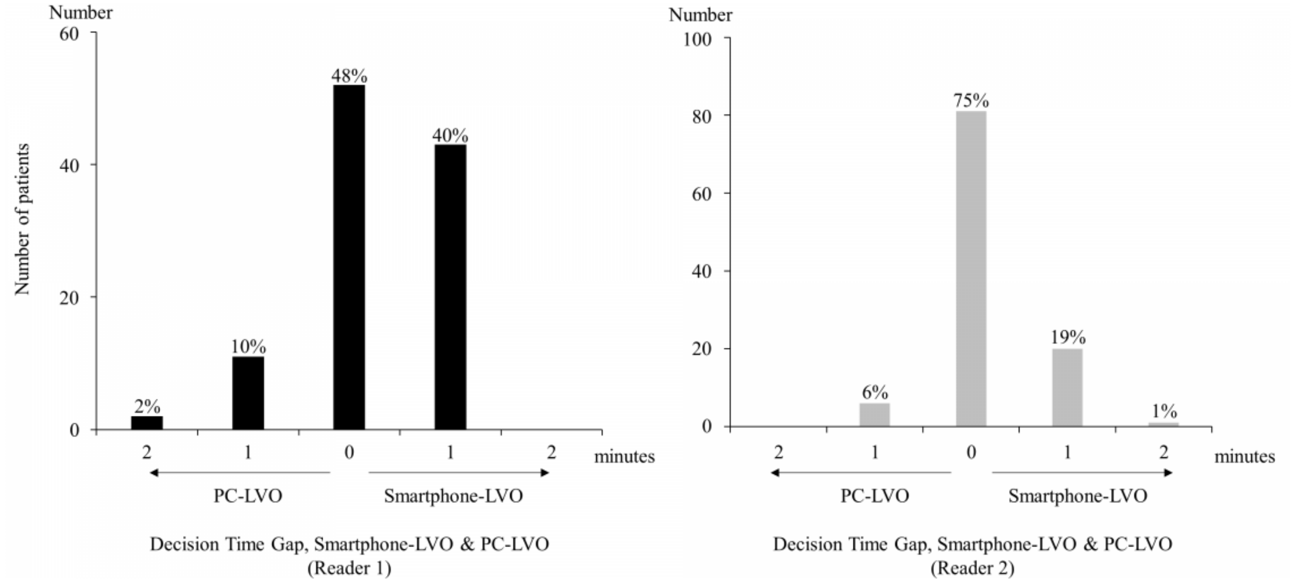
Wilcoxon signed-rank test (2-tailed) revealed that the interpretation time was significantly longer for Smartphone-LVO than PC-LVO (Reader 1: z=–3.547; *P*<.001 and Reader 2: z=–2.921; *P*=.003). However, the median interpretation time was the same between Smartphone-LVO and PC-LVO for each reader (Reader 1: 2 min for each method and Reader 2: 1 min for each method; Table 6). The interpretation time difference between Smartphone-LVO and PC-LVO was less than 1 min in 98% of the patient data sets in our series (Figure 4).

Table 6. Comparison of interpretation times between Smartphone-LVO and PC-LVO.

	Interpretation time of Smartphone-LVO[a]	Interpretation time of PC-LVO
Reader 1 (TK)		
Median minute (IQR)	2(2-2)	2(1-2)
Standard deviation	0.64	0.68
Min minute	1	1
Max minute	4	3
Reader 2 (KS)		
Median minute (IQR)	1(1-2)	1(1-1)
Standard deviation	0.49	0.34
Min minute	1	1
Max minute	3	2

[a] LVO: large vessel occlusion.

Figure 4. Difference in interpretation times between Smartphone-LVO and PC-LVO for Readers 1 and 2. The difference in interpretation time between Smartphone-LVO and PC-LVO was <1 min in 98% of all patient datasets. Reader 1, left; Reader 2, right. LVO: large vessel occlusion.



Discussion

Principal Findings

The study results show that the smartphone application yielded an accurate diagnosis of anterior intracranial arterial occlusion in times similar to diagnoses completed using the hospital-based desktop PC monitor. Several studies have reported the use of smartphone applications for the management of stroke [4,13-15]; however, no study has reported their use for assessing LVO. To the best of our knowledge, this is the first article to examine interdevice agreement and compare interpretation times for MRA and HVS on FLAIR imaging between a smartphone application and a hospital desktop PC monitor.

The JOIN smartphone application shows promise for use in acute stroke management for several reasons. JOIN provides a platform for simultaneously sharing important clinical information among the entire stroke team via PACS, whether on duty or on call. Because the information shared within the team is secured, JOIN can seamlessly support intrahospital teleradiology for MT in the “drip and ship” acute stroke management model. Conventionally, after receiving a stroke call, the vascular neurologist must immediately access a desktop PC and log into the network and server before browsing images sent via PACS. The neurologist’s interpretation of the neuroimaging examinations is then shared over the telephone with one person in the stroke team. JOIN can substantially improve the speed of this process by reducing the time taken to evaluate neuroimaging, particularly in situations where the neurologist is not near a desktop PC or is not at the hospital at the time that treatment for hyperacute stroke is being decided. The utilization of the JOIN application should simplify and shorten the MT decision-making process.

Regarding the telestroke system, while peer-to-peer network protocols are standard, they are limited in that the sender communicates only with one receiver in the stroke team, much like a closed circuit loop. In contrast, the JOIN application enables multicasting communications from one sender to many receivers. The team leader pre-authorizes which stroke team members participate in the secure group chat, and all of the patient’s data are centralized to the group chat. The entire stroke team can then receive notifications of the stroke patient’s arrival, view neuroimaging examinations, and review the medical plan via JOIN. Thus, important information is shared immediately among all members. JOIN also enables messaging among the team and displays neuroimaging files at a high resolution. Because the decision-making process is tracked via messages displayed on JOIN, the application plays a role in avoiding errors. Thus, combining the conventional telestroke system with JOIN would be desirable for arranging hyperacute stroke care in a complex and time-sensitive environment.

Finally, the telestroke system, comprising the 3 components of teleconsultation (doctor to patient), teleconferencing, and teleradiology, is confronted with developmental challenges in Japan. It is difficult for hospitals in Japan to introduce the real-time interactive video conference systems required for tele-examination using a hub and spoke model because there is no legal or regulatory provision for doctors on the hub side regarding patient examinations or a provision for medical decisions [16]. In addition, the Japanese government only approves medical practice via conventional encounters with a unique patient in emergency settings. Therefore, teleconsultation is only allowed under strict conditions (ie, for monitoring chronic disease after a second visit). In contrast, teleradiology is commonly used in Japan because of national insurance coverage. The advantages of using JOIN for acute stroke care include its portability, quick accessibility of target neuroimages, and its financial viability. In the near future, the Japanese Stroke Association plans to expand their practical guidelines for telestroke to include ICT solutions, under which JOIN would be considered for use.

Limitations

Our study has several limitations. First, the MRA images were evaluated by only 2 vascular neurologists, both of whom were experienced. In a future study, more evaluators should be used, including less experienced evaluators. Second, the study did not directly prove that the smartphone application was useful for early reperfusion therapy. Finally, the study only evaluated anterior circulation ischemic stroke; its applicability to posterior circulation ischemic stroke has yet to be investigated.

Conclusions

A smartphone application yielded an accurate diagnosis of anterior intracranial arterial occlusion in times similar to diagnoses completed using the hospital-based desktop PC monitor. It also enabled access to all patient information regardless of the user's location. Thus, this system can play an important role in organizing the stroke team to manage patients with hyperacute ischemic stroke.

The authors declare that all supporting data are available within the article.

Acknowledgments

None declared.

Authors' Contributions

TK contributed to the study design and concept, data acquisition and analysis, and drafted the manuscript for intellectual content. KS contributed substantially to the data acquisition. YI and YM revised the manuscript for intellectual content. HT and TI contributed to the data interpretation. YI approved the final version for publication.

Conflicts of Interest

HT and YM are inventors of the JOIN app. HT received private research funding from Allm, the developer of the JOIN app. The remaining authors declare no conflict of interest.

References

1. Fransen PSS, Berkhemer OA, Lingsma HF, Beumer D, van DBLA, Yoo AJ, Multicenter Randomized Clinical Trial of Endovascular Treatment of Acute Ischemic Stroke in the Netherlands Investigators. Time to Reperfusion and Treatment Effect for Acute Ischemic Stroke: A Randomized Clinical Trial. JAMA Neurol 2016 Feb;73(2):190-196. [doi: 10.1001/jamaneurol.2015.3886] [Medline: 26716735]
2. Saver JL, Goyal M, van der Lugt A, Menon BK, Majoie CBLM, Dippel DW, Chamorro, HERMES Collaborators. Time to Treatment With Endovascular Thrombectomy and Outcomes From Ischemic Stroke: A Meta-analysis. JAMA 2016 Oct 27;316(12):1279-1288. [doi: 10.1001/jama.2016.13647] [Medline: 27673305]
3. Saver JL. Time is brain--quantified. Stroke 2006 Jan;37(1):263-266. [doi: 10.1161/01.STR.0000196957.55928.ab] [Medline: 16339467]
4. Martins SC, Weiss G, Almeida AG, Brondani R, Carbonera LA, de Souza AC, et al. Validation of a Smartphone Application in the Evaluation and Treatment of Acute Stroke in a Comprehensive Stroke Center. Stroke 2020 Jan;51(1):240-246. [doi: 10.1161/strokeaha.119.026727]
5. 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]
6. Sawyer RN. Intravenous Tissue Plasminogen Activator for Large Vessel Ischemic Stroke - Is There Still a Role? Neurosurgery 2019 Jul 01;85(suppl_1):S34-S37. [doi: 10.1093/neuros/nyz083] [Medline: 31197340]
7. Campbell BCV, Donnan GA, Lees KR, Hacke W, Khatri P, Hill MD, et al. Endovascular stent thrombectomy: the new standard of care for large vessel ischaemic stroke. Lancet Neurol 2015 Aug;14(8):846-854. [doi: 10.1016/S1474-4422(15)00140-4] [Medline: 26119323]
8. 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;13(2):e31 [FREE Full text] [doi: 10.2196/jmir.1732] [Medline: 21550961]
9. 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 Jun 09;22(6):e15893 [FREE Full text] [doi: 10.2196/15893] [Medline: 32515744]
10. Adams HP, Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, et al. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. Stroke 1993 Jan;24(1):35-41. [doi: 10.1161/01.str.24.1.35] [Medline: 7678184]
11. 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]
12. Kamran S, Bates V, Bakshi R, Wright P, Kinkel W, Miletich R. Significance of hyperintense vessels on FLAIR MRI in acute stroke. Neurology 2000 Jul 25;55(2):265-269. [doi: 10.1212/wnl.55.2.265] [Medline: 10908902]

13. Shkirkova K, Akam EY, Huang J, Sheth SA, Nour M, Liang CW, UCLA Stroke Systems of Care Study Group. Feasibility and utility of an integrated medical imaging and informatics smartphone system for management of acute stroke. Int J Stroke 2017 Dec;12(9):953-960. [doi: 10.1177/1747493017694386] [Medline: 28358291]
14. Munich SA, Tan LA, Nogueira DM, Keigher KM, Chen M, Crowley RW, et al. Mobile Real-time Tracking of Acute Stroke Patients and Instant, Secure Inter-team Communication - the Join App. Neurointervention 2017 Sep;12(2):69-76 [FREE Full text] [doi: 10.5469/neuroint.2017.12.2.69] [Medline: 28955508]
15. Takao H, Murayama Y, Ishibashi T, Karagiozov KL, Abe T. A new support system using a mobile device (smartphone) for diagnostic image display and treatment of stroke. Stroke 2012 Jan;43(1):236-239 [FREE Full text] [doi: 10.1161/STROKEAHA.111.627943] [Medline: 21998052]
16. Schwamm LH, Audebert HJ, Amarenco P, Chumbler NR, Frankel MR, George MG, et al. Recommendations for the Implementation of Telemedicine Within Stroke Systems of Care. Stroke 2009 Jul;40(7):2635-2660. [doi: 10.1161/strokeaha.109.192361]

Abbreviations

FLAIR: fluid-attenuated inversion-recovery
FOV: field-of-view
HVS: hyperintense vessel sign
ICA: internal carotid artery
ICT: information and communication technology
IV-tPA: intravenous tissue plasminogen activator
LVO: large vessel occlusion
MRA: magnetic resonance angiography
MRI: magnetic resonance imaging
MT: mechanical thrombectomy
PACS: picture archiving and communication system
TE: echo time
TOF: time-of-flight
TR: repetition time

Edited by R Kukafka; submitted 24.02.21; peer-reviewed by A Roy, M Ghaderzadeh, J Marconato, RS Mahmoud, T Capote, S Chatterjee; comments to author 15.04.21; revised version received 21.05.21; accepted 14.06.21; published 27.08.21.

Please cite as:
Komatsu T, Sakai K, Iguchi Y, Takao H, Ishibashi T, Murayama Y
Using a Smartphone Application for the Accurate and Rapid Diagnosis of Acute Anterior Intracranial Arterial Occlusion: Usability Study
J Med Internet Res 2021;23(8):e28192
URL: https://www.jmir.org/2021/8/e28192
doi:10.2196/28192
PMID:34448716

©Teppei Komatsu, Kenichiro Sakai, Yasuyuki Iguchi, Hiroyuki Takao, Toshihiro Ishibashi, Yuichi Murayama. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

A Home-Based eHealth Intervention for an Older Adult Population With Food Insecurity: Feasibility and Acceptability Study

Luís Antunes Gomes[1,2], MSc; Maria João Gregório[1,2,3], PhD; Tatiana A Iakovleva[4], PhD; Rute Dinis de Sousa[1,2], MSc, MBA; John Bessant[5], PhD; Pedro Oliveira[6,7], PhD; Jaime C Branco[1,2,8], MD, PhD; Helena Canhão[1,2], MD, PhD; Ana Maria Rodrigues[1,2], MD, PhD

[1]Comprehensive Health Research Centre, NOVA Medical School, Universidade NOVA de Lisboa, Lisbon, Portugal
[2]EpiDoC Unit, Centro de Estudos de Doenças Crónicas, NOVA Medical School, Universidade NOVA de Lisboa, Lisbon, Portugal
[3]Faculdade de Ciências da Nutrição e Alimentação da Universidade do Porto, Porto, Portugal
[4]Stavanger Business School, Stavanger, Norway
[5]University of Exeter Business School, Exeter, United Kingdom
[6]Copenhagen Business School, Copenhagen, Denmark
[7]NOVA School of Business and Economics, Universidade NOVA de Lisboa, Lisbon, Portugal
[8]Serviço de Reumatologia do Hospital Egas Moniz - Centro Hospitalar Lisboa Ocidental (CHLO-EPE), Lisbon, Portugal

Corresponding Author:
Ana Maria Rodrigues, MD, PhD
Comprehensive Health Research Centre
NOVA Medical School
Universidade NOVA de Lisboa
Pólo de Investigação, NMS, UNL. Edifício Amarelo
Piso 2. Rua do Instituto Bacteriológico, nº5
Lisbon
Portugal
Phone: 351 218 803 000
Email: ana.m.rodrigues@nms.unl.pt

Abstract

Background: Food insecurity is a global public health challenge, affecting predominately the most vulnerable people in society, including older adults. For this population, eHealth interventions represent an opportunity for promoting healthy lifestyle habits, thus mitigating the consequences of food insecurity. However, before their widespread dissemination, it is essential to evaluate the feasibility and acceptability of these interventions among end users.

Objective: This study aims to explore the feasibility and acceptability of a home-based eHealth intervention focused on improving dietary and physical activity through an interactive television (TV) app among older adults with food insecurity.

Methods: A pilot noncontrolled quasi-experimental study was designed with baseline and 3-month follow-up assessments. Older adult participants with food insecurity were recruited from 17 primary health care centers in Portugal. A home-based intervention program using an interactive TV app aimed at promoting healthy lifestyle behaviors was implemented over 12 weeks. Primary outcomes were feasibility (self-reported use and interest in eHealth) and acceptability (affective attitude, burden, ethicality, perceived effectiveness, and self-efficacy), which were evaluated using a structured questionnaire with a 7-point Likert scale. Secondary outcomes were changes in food insecurity (Household Food Insecurity Scale), quality of life (European Quality of Life Questionnaire with five dimensions and three levels and Functional Assessment of Chronic Illness Therapy-Fatigue), physical function (Health Assessment Questionnaire, Elderly Mobility Scale, grip strength, and regularity of exercise), and nutritional status (adherence to the Mediterranean diet).

Results: A sample of 31 older adult individuals with food insecurity was enrolled in the 12-week intervention program with no dropouts. A total of 10 participants self-reported low use of the TV app. After the intervention, participants were significantly more interested in using eHealth to improve food insecurity (baseline median 1.0, IQR 3.0; 3-month median 5.0, IQR 5.0; *P*=.01) and for other purposes (baseline median 1.0, IQR 2.0; 3-month median 6.0, IQR 2.0; *P*=.03). High levels of acceptability were found both before and after (median range 7.0-7.0, IQR 2.0-0.0 and 5.0-7.0, IQR 2.0-2.0, respectively) the intervention, with no significant changes for most constructs. Clinically, there was a reduction of 40% in food insecurity (*P*=.001), decreased fatigue (mean −3.82, SD 8.27; *P*=.02), and improved physical function (Health Assessment Questionnaire: mean −0.22, SD 0.38; *P*=.01; Elderly Mobility Scale: mean −1.50, SD 1.08; *P*=.01; regularity of exercise: baseline 10/31, 32%; 3 months 18/31, 58%; *P*=.02). No differences were found for the European Quality of Life Questionnaire with five dimensions and three levels, grip strength, or adherence to the Mediterranean diet.

Conclusions: The home-based eHealth intervention was feasible and highly acceptable by participants, thus supporting a future full-scale trial. The intervention program not only reduced the proportion of older adults with food insecurity but also improved participants' fatigue and physical function.

International Registered Report Identifier (IRRID): RR2-10.2196/resprot.6626

(J Med Internet Res 2021;23(8):e26871) doi:10.2196/26871

KEYWORDS

food insecurity; eHealth; television app; elderly people; vulnerable population; cognitive behavioral strategy; health innovation; multidisciplinary program

Introduction

Background

Food security is an essential prerequisite for a population to be healthy, active, and well-nourished. According to the World Health Organization, food security exists "when all people, at all times, have physical, social and economic access to sufficient, safe and nutritious food that meets their dietary needs and food preferences for an active and healthy life" [1]. As a comprehensive concept, food security depends on four essential dimensions: physical availability of food, economic and physical access to food, food use and maximization of consumption, and stability at all times [1]. When one or more of these dimensions are compromised, people or households are assumed to have food insecurity.

The multidimensional and self-perceived nature of food insecurity makes quantification challenging. Among the various instruments, the Household Food Insecurity Scale developed by the United States Department of Agriculture (USDA) is the most widely used tool in epidemiological studies [2,3]. This instrument classifies the food insecurity status of an entire household while considering multiple dimensions using a simple, easy, and validated approach [3].

Food insecurity is a global public health problem. Despite mainly affecting low-income countries, increasing evidence suggests that food insecurity is also highly prevalent in high-income countries [4]. For example, in the United States, one of the richest nations in the world, a USDA report states that 10.5% of households were facing food insecurity in 2019 [2]. Although this represents a lower prevalence since the Great Recession of 2008, the economic downturn caused by the COVID-19 pandemic may trigger an unprecedented food insecurity crisis [5,6]. National estimates in March and April 2020 indicate that the prevalence of this problem has more than tripled to 38%, which is the highest level of food insecurity ever measured in the United States [7].

In Europe, particularly Portugal, a country still recovering from an economic crisis, the impact of the COVID-19 pandemic is uncertain. The current economic, political, and social instability in this country may have overwhelming consequences, especially for individuals who are physically, economically, and socially vulnerable, such as older adults. Indeed, the prevalence of food insecurity among older adults is remarkably high, with a study from our group reporting that 23% of older Portuguese people experienced food insecurity in 2015 and 2016, which is higher than the prevalence among adults [8,9].

The consequences of food insecurity among older adults present major challenges to the society. The aging of this population, which has a poor overall health status, physical comorbidities, limited family contact or assistance, and low income, can compromise the motivation of older adults to adopt health behaviors that are compatible with an active and healthy lifestyle, particularly those related to a healthy diet and physical activity. For this population, which has special nutritional needs, food insecurity is associated with a poor nutritional status, physical disability, muscle weakness, depression and anxiety symptoms, and lower health-related quality of life (HRQoL), which in turn increases vulnerability to other conditions, hospitalization, and death [9-14].

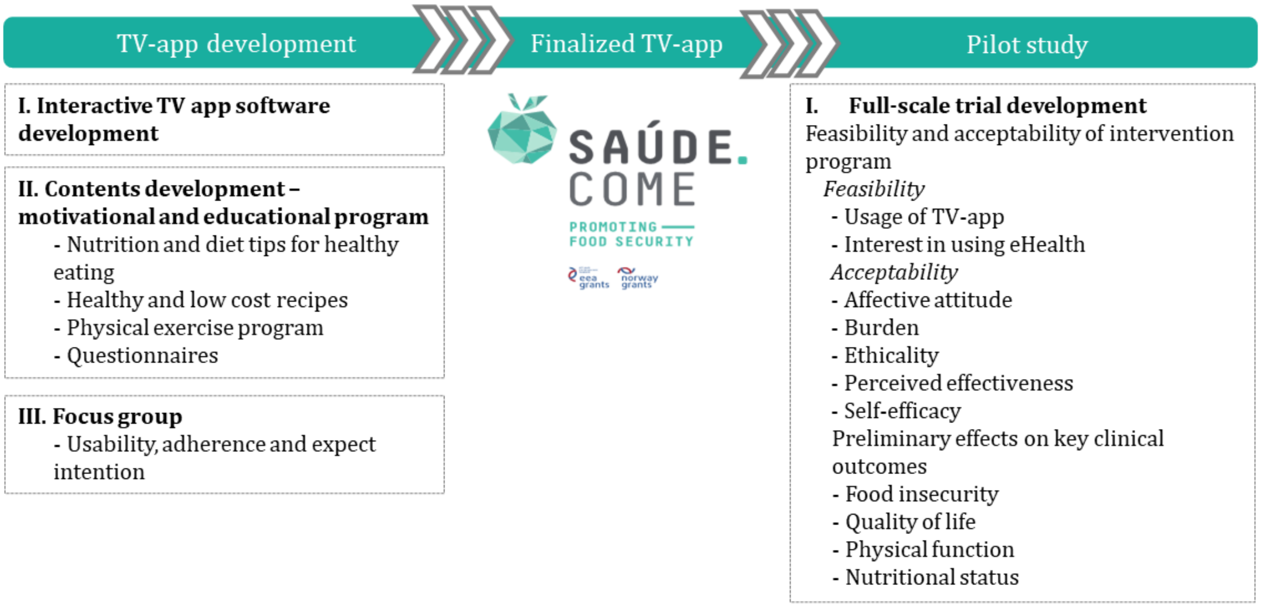
Developing strategies to mitigate the consequences of food insecurity among older adults is essential for increasing their quality of life and well-being. However, the unique characteristics of this population make it difficult to conduct an experimental study, leaving this particularly vulnerable and growing population with low access to health care innovation.

The promotion of healthy lifestyle habits and the resulting increase in well-being and quality of life among older adults could be addressed by eHealth interventions, which are defined as the use of information and communication technologies (ICTs) for health [15]. These interventions are low-cost, personalized approaches capable of increasing the autonomy of older adults and their access to quality health care [16-18]. Emerging evidence shows that eHealth interventions for older adults have delivered promising results in terms of improving nutritional status [19] and physical activity [20,21] and have positive clinical outcomes for specific conditions such as fragility fracture [22] and cardiovascular disease [23]. However, unique characteristics of the lifestyle of older adults should be acknowledged to overcome the associated barriers before the wider implementations of eHealth interventions [18,19,24].

Among the available ICTs, television (TV) is the device mostly used by older people [25]. As most older people's houses have a TV, this device could present an opportunity to successfully design and implement eHealth interventions aimed at promoting healthy lifestyles with their daily life. By improving the nutritional status and physical activity levels, a TV home-based program might slow the decline in physical function and increase the well-being and quality of life among older adults experiencing food insecurity.

To promote healthy lifestyles among older adults with food insecurity, we designed a home-based intervention program using an interactive TV app [26]. Following the new Medical Research Council's guidance for developing and evaluating complex interventions [27], a multidisciplinary team of health professionals (ie, physicians, nurses, nutritionists, and physiotherapists), physical exercise experts, telecommunications companies, stakeholders, and end users collaboratively developed the conceptual framework of the program (Figure 1). In this development process, 11 older adult participants in a focus group reported high app usability, adherence intention, and the expected impact of the program on behavior modification [26]. All aspects of program conception considered ethical, social, economic, cultural, and environmental implications for all stakeholders, as suggested by the Responsible Innovation Framework [28].

Figure 1. Conceptual framework of the program. TV: television.


TV-app development
Finalized TV-app
Pilot study
I. Interactive TV app software development
II. Contents development – motivational and educational program
- Nutrition and diet tips for healthy eating
- Healthy and low cost recipes
- Physical exercise program
- Questionnaires
III. Focus group
- Usability, adherence and expect intention
SAÚDE. COME
PROMOTING FOOD SECURITY
eea grants
norway grants
I. Full-scale trial development
Feasibility and acceptability of intervention program
Feasibility
- Usage of TV-app
- Interest in using eHealth
Acceptability
- Affective attitude
- Burden
- Ethicality
- Perceived effectiveness
- Self-efficacy
Preliminary effects on key clinical outcomes
- Food insecurity
- Quality of life
- Physical function
- Nutritional status

Objectives

After the conclusion of the TV app development process [26], the next step was to analyze the feasibility and acceptability of this intervention program in a pilot study. The primary aim of this pilot study is to test the feasibility and acceptability of a multidisciplinary 12-week home-based intervention program focused on improving dietary and physical activities through an interactive TV app among older adults with food insecurity. The secondary aim is to assess changes in the key clinical outcomes at the 3-month follow-up. The results of this pilot study are crucial for understanding end users' interest and use of the TV app along with the suitability of the program to their needs.

Methods

Research Design

This pilot study used a noncontrolled quasi-experimental design with before and after measurements. Program feasibility, acceptability, and key clinical outcomes were evaluated at baseline and at the 3-month follow-up after the completion of the eHealth intervention. Older adult participants aged 60 years or above were recruited from 17 primary health care centers in the *Lisboa e Vale do Tejo* health region in Portugal.

Participants

A convenience sample of older adults with food insecurity from 17 primary care centers in *Lisboa e Vale do Tejo* health region was used in this study. Individuals were included if they (1) were aged 60 years or above, (2) were classified as having food insecurity during recruitment [29], (3) were able and willing to give written informed consent and a phone number contact and comply with other requirements of the study protocol, (4) were a Portuguese speaker or able to understand Portuguese, (5) were noninstitutionalized and living in a private household in Portugal, (6) had electricity at home, and (7) had cable TV with a box at home. Individuals were excluded if they (1) were unable to comply with the study protocol (eg, hearing or visual loss and cognitive impairment), (2) had an absolute contraindication to exercise, (3) had cable TV operated by a company other than MEO, Vodafone, or NOS, and (4) had a household member involved in the study. Participants were recruited between November 2015 and May 2016. During the recruitment phase, all older adults who attended health services in the selected primary health care centers were invited to complete an initial screening questionnaire, which included a food insecurity scale [29]. Participants who agreed to complete the questionnaire were asked to provide their consent for further contact. All participants who identified as having food insecurity were invited to participate in this study. In case of acceptance, the first baseline assessment appointment was scheduled. This occurred during an appointment in a primary health care setting with a multidisciplinary team that included a physician, nurse, nutritionist, and physiotherapist. The physician asked for informed consent and checked the eligibility criteria.

Intervention

A detailed description of the intervention can be found in a previous study [26], which briefly comprised a 12-week home-based intervention focused on the innovative use of an interactive TV app. The design of the intervention program was based on a transtheoretical model of behavior change [30] and aimed to promote lifestyle behavior changes among older individuals with food insecurity by (1) providing education on the importance of a healthy diet and physical activity among older adults, (2) demonstrating that low household income is not a barrier to healthy lifestyle habits and that it is possible to have a healthy diet and practice physical activity at low cost, and (3) providing motivation for adopting a healthy diet and physical activity habits to reduce noncommunicable diseases. The pillars of program content were divided into nutrition and diet tips for healthy eating, low-cost healthy recipes, and physical exercise programs. Different contents were specifically structured based on the thematic weeks. The themes for each week and the respective content of each video were developed by considering the different types of food most frequently used in daily life routine (eg, vegetables, water, and milk). Nutrition and diet tips were based on the explanations of the benefits and harms of eating certain foods, especially regarding their relationship with risk factors for noncommunicable diseases and nutritional requirements for older adults [31-34]. Low-cost healthy recipes were specifically designed for our program using a popular Portuguese TV chef and a nutritionist. Finally, the physical activity program was developed by physical exercise experts with the specific objective of promoting the practice of physical activity at home for at least 30 minutes thrice a week [35].

All program content was disseminated on a dedicated TV channel via the interactive TV app and delivered on a scheduled basis on specific days of the week for developing a program access routine. To increase the participants' motivation, interactive TV reminders, including tips about healthy lifestyle habits, were sent on a weekly basis. In addition, participants had at their disposal the teams' telephone numbers and options to contact them whenever necessary. The research team also made frequent contact with the participants to increase their adherence and overcome difficulties. The TV app software included the delivery of small questionnaires aimed at assessing participants' compliance and learning during the intervention program. These questionnaires were delivered on a weekly basis and could be answered using TV remote control buttons. Any participant who opened the TV app fewer than 2 times during the intervention period was excluded from the data analysis.

Outcomes

Overview

The primary outcomes were feasibility and acceptability of the intervention program. Secondary outcomes were changes in food insecurity, quality of life, physical function, and nutritional status after the intervention.

Data were collected using a structured questionnaire administered at two time points: before the intervention at study enrollment (baseline) and 3 months after the intervention (follow-up). The baseline assessment was performed by a multidisciplinary team (ie, physicians, nurses, nutritionists, and physiotherapists) at the primary care centers involved in the study using a computer-assisted personal interview system. At the 3-month follow-up, the assessment was performed by telephone using a computer-assisted personal interview system by a team of trained research assistants.

Baseline Characteristics

During baseline assessment, information was collected regarding sociodemographic (ie, gender, age, years of education, and marital status), socioeconomic (ie, employment status, household composition, household monthly income, and income perception), clinical (ie, food insecurity, BMI [kg/m^2], anxiety and depression symptoms measured using the Hospital Anxiety and Depression Scale [36], and self-reported noncommunicable chronic diseases), and lifestyle characteristics (ie, alcohol intake profile, smoking habits, PREDIMED [Prevención con Dieta Mediterránea] score, perceptions of the importance and difficulty of healthy eating, frequency of watching TV, frequency of computer, videogame, tablet use, and frequency of internet use).

Feasibility and Acceptability of the Intervention Program

The measures of feasibility of the intervention program were (1) self-reported use of the TV app evaluated at 3-month follow-up and (2) interest in the use of eHealth evaluated with a structured questionnaire administered pre- (baseline) and postintervention (3-month follow-up). This questionnaire was also used to evaluate the acceptability of the intervention program. The acceptability evaluation was based on the theoretical framework of acceptability and included the following component constructs: affective attitude ("how an individual feels about the intervention"), burden ("perceived amount of effort required to participate in the intervention"), ethicality ("extent to which the intervention has good fit with an individuals' value system"), perceived effectiveness ("extent to which the intervention is perceived as likely to achieve its purpose"), and self-efficacy ("individuals' confidence that they can perform the behaviors required to participate in the intervention") [37]. The structured questionnaire included 16 questions rated on a 7-point Likert scale with scores ranging from 1 (strongly negative/completely disagree) to 7 (strongly positive/completely agree). To understand participants' perceptions in greater detail, the questionnaire included 3 additional questions using an open question approach to evaluate the most liked and disliked content of the intervention program (affective attitude construct) and the greatest difficulties in using the TV app (burden construct). Participants who initially self-reported low use of the TV app were not included in the evaluation of the interest construct of feasibility or the acceptability of the intervention program, as their experience with the TV app was limited. The reasons for the low use of the TV app were assessed using an open question approach.

Food Insecurity, Quality of Life, Physical Function, and Nutritional Status

Secondary outcomes were changes in food insecurity, quality of life, physical function, and nutritional status after the intervention. Food insecurity was measured using the Household Food Insecurity Scale, a scale adapted and validated for the Portuguese population from the USDA Household Food Security Survey Module [29]. The tool is applied at the individual level and collects data on food insecurity status for the whole household. Using a score ranging from 0 to 14, households were classified into different categories of food insecurity: food security (score of 0), low food insecurity (score between 1 and 5 for households with children and between 1 and 3 for households without children), moderate food insecurity (score between 6 and 9 for households with children and between 4 and 5 for households without children), and severe food insecurity (score between 10 and 14 for households with children and between 6 and 8 for households without children) [29].

Quality of life was evaluated using two different measures. The European Quality of Life Questionnaire with five dimensions and three levels (EQ-5D-3L) was used to measure HRQoL [38,39]. A higher EQ-5D-3L score corresponded to a higher quality of life. The Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F) was used to evaluate fatigue in patients with chronic diseases [40]. This score ranges from 0 to 52, with higher scores representing less fatigue.

The physical function was evaluated using several instruments. The functional ability of the individuals was measured using the Health Assessment Questionnaire (HAQ) [41,42]. The final HAQ score ranges from 0 to 3 points, with higher scores corresponding to a lower functional ability. The Elderly Mobility Scale (EMS) was used to evaluate the performance and mobility of participants and has a score ranging from 0 to 20 [43,44], with higher scores indicating a greater degree of independence or mobility. A dynamometer was used to measure hand grip strength (lbs). Finally, data concerning exercise regularity (yes or no) and frequency (0, 1-2, 2-4, or ≥5 exercise sessions per week for a duration of at least 45 minutes) were recorded.

Nutritional status was classified based on the PREDIMED questionnaire, a 14-item questionnaire assessing adherence to the Mediterranean diet (MD) [8]. A score of ≥10 points indicated a high adherence to MD, and a score <10 points indicated a low adherence to MD [45].

Statistical Analysis

The results were analyzed using SPSS Statistics (version 26.0, IBM Corp). Continuous variables are reported as mean and SD. Categorical variables are reported as frequencies or proportions. Changes in the feasibility and acceptability of the intervention program between baseline and 3-month follow-up are reported as median and IQR. The Wilcoxon signed-rank test or Sign test was used to assess differences after the intervention regardless of whether the data were symmetric. Differences in food insecurity, quality of life, physical function, and nutritional status between baseline and 3-month follow-up were assessed depending on the type of variable (continuous or ordinal), normal distribution, and data symmetry. The food insecurity results are summarized as frequencies and proportions in a contingency table. The Sign test was used to assess the differences between pre- and postintervention, as all participants had food insecurity at baseline. EQ-5D-3L, FACIT-F, HAQ, EMS, and hand grip strength data are reported as mean and SD, and differences at 3-month follow-up were evaluated using a 2-tailed paired *t* test, Wilcoxon signed-rank test, or Sign test. Regularity of exercise (yes or no) and PREDIMED (low or high adherence to MD) results were analyzed using frequencies and proportions in a contingency table, and differences were evaluated using McNemar tests for paired binary data. A significance level of 5% was considered as statistically significant.

Ethical Issues

The study was performed in accordance with the principles established by the Declaration of Helsinki and was reviewed and approved by the NOVA Medical School Ethics Committee, the National Committee for Data Protection (*Comissão Nacional de Proteção de Dados*), and the Ethical Committee of the Regional Health Authority of *Lisboa e Vale do Tejo*.

Results

Baseline Characteristics

During the 7-month recruitment period, 1857 individuals were screened at primary health care centers, 628 of whom were classified as having food insecurity. Of these, 177 older adult participants were identified as potentially eligible for the study; however, most did not agree to participate (73/177, 41.2%) or missed the medical appointment (45/177, 25.4%). Several participants (21/177, 11.9%) were excluded because of technological incompatibilities with the interactive TV app installation or software (Figure 2). Table 1 presents the participants' baseline sociodemographic, socioeconomic, clinical, and lifestyle characteristics (n=31). The participants were predominantly female (21/31, 68%) with a mean age of 71.9 (SD 6.6) years and a low education profile (10/23, 43%). Most participants were retired (24/30, 80%), lived in a household with 3 or more people (9/23, 39%), and had a very low monthly income (13/23, 57%). Clinically, 50% (11/22) of the participants were obese, and most had chronic noncommunicable diseases. Regarding lifestyle, only 32% (10/31) were physically active and 90% (19/21) reported low adherence to MD. All participants acknowledged the importance of healthy eating (22/22, 100%), despite recognizing the associated difficulties (10/22, 45%). Finally, with a predominant watching frequency of 2-3 hours per day (13/31, 42%), TV was the predominantly used ICTs by participants. No participants were excluded from the data analysis at the 3-month follow-up.

Figure 2. Participant recruitment and retention flowchart. TV: television.

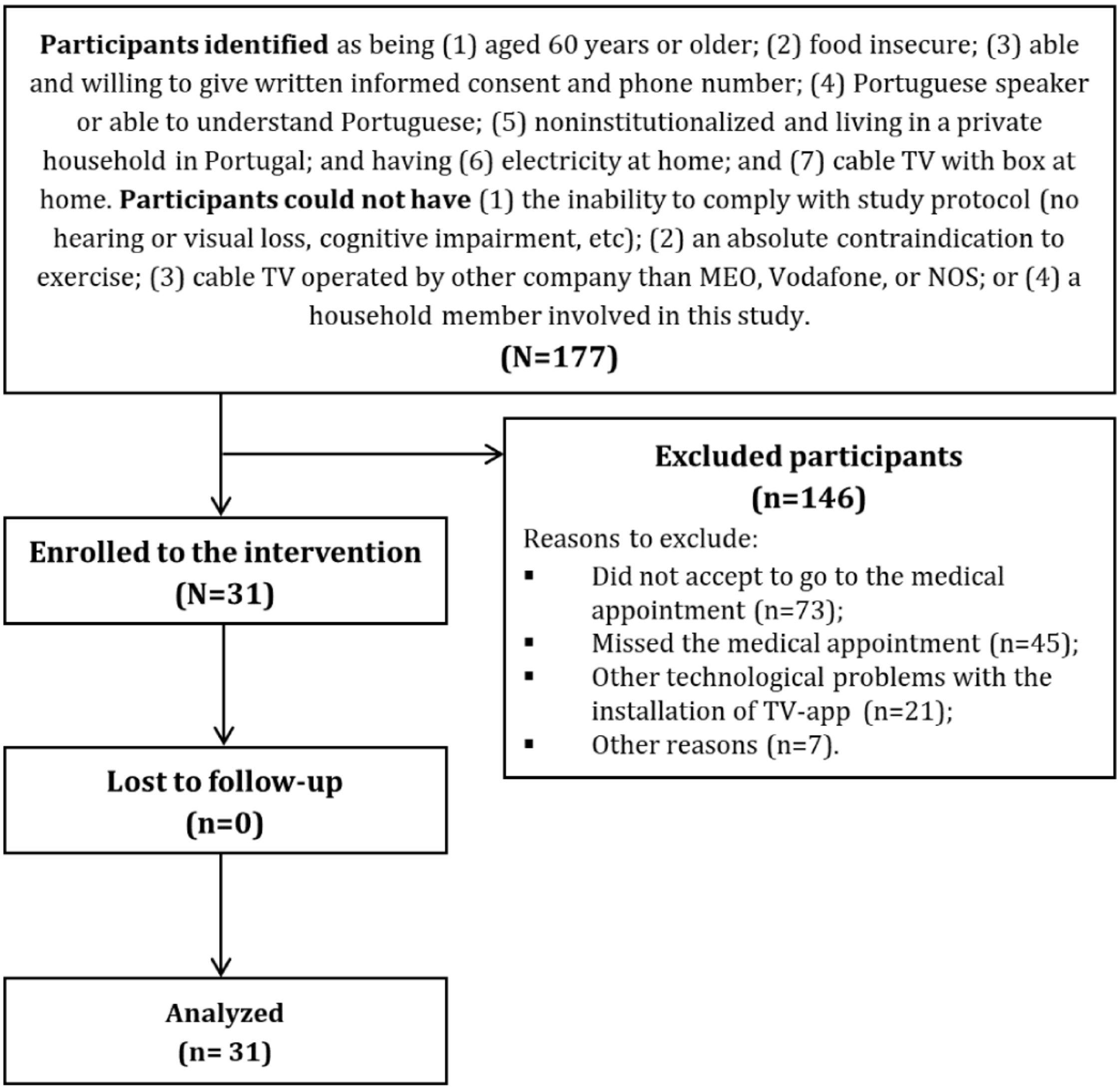

Participants identified as being (1) aged 60 years or older; (2) food insecure; (3) able and willing to give written informed consent and phone number; (4) Portuguese speaker or able to understand Portuguese; (5) noninstitutionalized and living in a private household in Portugal; and having (6) electricity at home; and (7) cable TV with box at home. **Participants could not have** (1) the inability to comply with study protocol (no hearing or visual loss, cognitive impairment, etc); (2) an absolute contraindication to exercise; (3) cable TV operated by other company than MEO, Vodafone, or NOS; or (4) a household member involved in this study.
(N=177)
Excluded participants
(n=146)
Reasons to exclude:
Did not accept to go to the medical appointment (n=73);
Missed the medical appointment (n=45);
Other technological problems with the installation of TV-app (n=21);
Other reasons (n=7).
Enrolled to the intervention
(N=31)
Lost to follow-up
(n=0)
Analyzed
(n= 31)

Table 1. Baseline sociodemographic, socioeconomic, clinical, and lifestyle characteristics of participants (N=31)[a].

Characteristics	Values
Sociodemographic	
Gender (n=31), n (%)	
Female	21 (68)
Age (years; n=31), mean (SD)	71.9 (6.6)
Age group (years; n=31), n (%)	
60-69	15 (48)
70-79	10 (32)
80	6 (19)
Years of education (n=23), n (%)	
0-4	10 (43)
5-9	6 (26)
10	7 (30)
Marital status (n=24), n (%)	
Single	3 (13)
Married	13 (54)
Divorced	6 (25)
Widow or widower	2 (8)
Socioeconomic, n (%)	
Employment status (n=30)	
Employed	1 (3)
Unemployed	5 (17)
Retired	24 (80)
Household composition (n=23)	
1 person	8 (35)
2 people	6 (26)
≥3 people	9 (39)
Household monthly income (€; US $; n=23)	
750 (880.63)	13 (57)
751-1000 (881.80-1174.17)	5 (22)
>1000 (1174.17)	5 (22)
Income perception (n=23)	
I live comfortably with my current income	1 (4)
I can live with my current income	7 (30)
It is difficult to live with my current income	7 (30)
It is very difficult to live with my current income	8 (35)
Clinical, n (%)	
Food insecurity (n=31)	
Food security	0 (0)
Food insecurity	31 (100)
Low food insecurity	27 (87)
Moderate food insecurity	3 (10)
Severe food insecurity	1 (3)

Characteristics	Values
BMI (kg/m^2; n=22)	
Underweight	0 (0)
Normal weight	3 (14)
Overweight	8 (36)
Obesity	11 (50)
Anxiety and depression symptoms (n=23)	
Anxiety symptoms (HADS[b] score ≥11)	7 (30)
Depression symptoms (HADS score ≥11)	3 (13)
Noncommunicable chronic diseases (self-reported; n=31)	
Rheumatic disease	20 (65)
High blood pressure	17 (55)
Diabetes	7 (23)
High cholesterol	12 (39)
Pulmonary disease	2 (6)
Cardiac disease	5 (16)
Gastrointestinal disease	3 (10)
Neurologic disease	4 (13)
Neoplastic disease	3 (10)
Thyroid or parathyroid disease	5 (16)
Other diseases (eg, ophthalmic, dermatologic, or auditory)	3 (10)
Lifestyle habits, n (%)	
Physical activity (n=31)	
Regular	10 (32)
Alcohol intake profile (n=23)	
Daily	2 (9)
Occasionally	12 (52)
Never	9 (39)
Smoking habits (n=31)	
Current smoker	4 (13)
Past smoker	9 (29)
Never	18 (58)
PREDIMED[c] (n=21)	
Low adherence to MD[d]	19 (90)
High adherence to MD	2 (10)
Perceptions about the importance of healthy eating (n=22)	
Without or with little importance	0 (0)
Neither important nor unimportant	0 (0)
Important or very important	22 (100)
Perceptions about the difficulty of healthy eating (n=22)	
Difficult or very difficult	10 (45)
Neither difficult nor easy	3 (14)
Easy or very easy	9 (41)

Characteristics	Values
Frequency of watching TV[e] (n=31)	
Does not watch	1 (3)
1 hours/day	6 (19)
2-3 hours/day	13 (42)
≥4 hours/day	11 (35)
Frequency of computer, videogame, or tablet use (n=31)	
Does not use	16 (52)
1 hours/day	6 (19)
2-3 hours/day	7 (23)
≥4 hours/day	2 (6)
Frequency of internet use (n=31)	
Does not use	16 (52)
1 hours/day	8 (26)
2-3 hours/day	6 (19)
≥4 hours/day	1 (3)

[a] Sample size is not constant due to missing data.

[b] HADS: Hospital Anxiety and Depression Scale.

[c] PREDIMED: Prevención con Dieta Mediterránea.

[d] MD: Mediterranean diet.

[e] TV: television.

Feasibility and Acceptability of the Intervention Program

After the intervention program, some participants (n=10) self-reported low use of the TV app, and thus did not provide answers to the feasibility construct “interest in using eHealth” or acceptability questions. The reasons for the low use of the TV app were “lack of interest” (n=2), “lack of time” (n=2), “much time away from home during the intervention period” (n=2), “complications with the software” (n=1), or “health reasons” (n=1). Two participants did not provide a reason. Despite these results, all participants opened the TV app two or more times, and none were excluded from the data analysis of clinical outcome.

Regarding participants’ interest in using eHealth, a considerable increase was observed both for the treatment of food insecurity and other purposes.

Acceptability constructs showed very high scores both at baseline and at the 3-month follow-up, with no significant changes for most items (Table 2). However, “interest in using the TV app in daily life,” “attractiveness of using the TV app in daily life,” “advantages of using the TV app are/were higher than disadvantages,” and “ability of the TV app to solve challenges/problems in daily life” were significantly lower after the intervention program. The most liked contents were exercise (12/20, 60%), cooking (12/20, 60%), and nutrition (10/20, 50%) whereas the most dislike were exercise (5/20, 25%) and questionnaires (2/20, 10%). The main difficulties on using the TV app were “difficult in the access or manage the app” (6/20, 30%) and “lack of time” (4/20, 20%).

Table 2. Acceptability and feasibility of the television app intervention program (n=21)[a].

Characteristics	Baseline[b], median (IQR)	3-month follow-up[b], median (IQR)	Difference[b], median (IQR)	*P* value
Feasibility				
Interest in using eHealth for treatment of food insecurity (n=20)	1.0 (3.0)	5.0 (5.0)	3.0 (5.0)	.01[c]
Interest in using eHealth for other purposes (n=20)	1.0 (2.0)	6.0 (2.0)	3.0 (4.0)	.03[c]
Acceptability				
Affective attitude				
Desire to use TV[d] app in daily life (n=19)	7.0 (1.0)	6.0 (2.0)	0.0 (2.0)	.06
Interest in using TV app in daily life (n=20)	7.0 (1.0)	6.0 (2.0)	0.0 (2.0)	.02[c]
Attractiveness of using the TV app in daily life (n=19)	7.0 (0.0)	6.0 (3.0)	0.0 (1.0)	.04[c]
Advantages of using the TV app are or were higher than disadvantages (n=19)	7.0 (0.0)	6.0 (3.0)	−1.0 (3.0)	.01[c]
Burden				
Ease of using the TV app (n=21)	7.0 (1.0)	6.0 (3.0)	−1.0 (1.0)	.09
Ethicality				
Ethical appropriateness of using digital technologies for treatment (n=20)	7.0 (0.0)	6.0 (1.0)	−1.0 (1.0)	.08
Acceptability of not including person-to-person interaction in medical treatment (n=20)	7.0 (2.0)	6.0 (2.0)	−1.0 (2.0)	.30
Suitability of using digital products to solve health problems (n=18)	7.0 (2.0)	5.0 (2.0)	−2.0 (3.0)	.08
Generally in favor of using technologies to treat people (n=20)	7.0 (0.0)	6.0 (2.0)	0.0 (2.0)	.14
Perceived effectiveness				
Utility of TV app (n=19)	7.0 (0.0)	6.0 (2.0)	0.0 (2.0)	.10
Ability of TV app to solve challenges or problems in daily life (n=16)	7.0 (0.0)	5.0 (1.0)	−2.0 (3.0)	.002[c]
Self-efficacy				
Motivation to use digital products (n=20)	7.0 (1.0)	7.0 (2.0)	0.0 (0.75)	.17
Control over TV app (n=20)	7.0 (1.0)	7.0 (2.0)	0.0 (1.5)	.47
Existence of circumstances beyond control that prevent use of TV app (n=20)	7.0 (1.0)	6.5 (4.0)	0.0 (2.5)	.05

[a] Sample size is not constant due to missing data.

[b] Items rated on a 7-point Likert scale from 1 (strongly negative/completely disagree) to 7 (strongly positive/completely agree).

[c] *P*<.05.

[d] TV: television.

Food Insecurity, Quality of Life, Physical Function, and Nutritional Status

Table 3 presents the changes in food insecurity, quality of life, physical function, and nutritional status. The intervention program significantly reduced food insecurity status and severity at the 3-month follow-up, at which point most participants who maintained food insecurity were classified into the low food insecurity subgroup. Regarding quality of life, the intervention had no impact on EQ-5D-3L, but significantly improved the FACIT-F scores. Regarding physical function, there was a significant improvement in HAQ, EMS, and regularity of exercise practice. No differences were found in hand grip strength after the intervention. Regarding nutritional status, adherence to MD did not change.

Table 3. Effects of the intervention program on food insecurity, quality of life, physical function, and nutritional status (N=31)[a].

Characteristics	Baseline	3-month follow-up	Difference	*P* value
Food security (n=30), n (%)	0 (0)	12 (40)	12 (40)	.001[b]
Food insecurity, n (%)	30 (100)	18 (60)	−12 (40)	N/A[c]
Low food insecurity	26 (87)	17 (57)	−9 (30)	
Moderate food insecurity	3 (10)	0 (0)	−3 (10)	
Severe food insecurity	1 (3)	1 (3)	0 (0)	
Quality of life, mean (SD)				
EQ-5D-3L[d] (n=23)	0.62 (0.29)	0.65 (0.30)	0.03 (0.26)	.58
FACIT-F[e] (n=31)	38.52 (9.32)	41.96 (10.00)	3.82 (8.27)	.02[b]
Physical function				
HAQ[f] (n=23), mean (SD)	0.77 (0.70)	0.55 (0.60)	−0.22 (0.38)	.01[b]
EMS[g] (n=10), mean (SD)	19.70 (0.67)	18.20 (1.23)	−1.50 (1.08)	.01[b]
Hand grip strength (n=10), mean (SD)	29.35 (7.52)	30.87 (10.44)	1.52 (3.72)	.21
Regular exercise (n=31), n (%)				
Yes	10 (32)	18 (58)	8 (26)	.02[b]
Days physically active in the last week (45-minutes duration; n=31), n (%)				N/A
None	24 (77)	16 (52)	−8 (26)	
1-2	3 (10)	8 (26)	5 (16)	
3-4	1 (3)	4 (13)	3 (10)	
5	3 (10)	3 (10)	0 (0)	
Nutritional status				
PREDIMED[h] (n=21), n (%)				.99
Low adherence to MD[i]	19 (90)	19 (90)	0 (0)	
High adherence to MD	2 (10)	2 (10)	0 (0)	

[a] Sample size is not constant due to missing data.

[b] *P*<.05.

[c] N/A: not applicable.

[d] EQ-5D-3L: European quality of life questionnaire with five dimensions and three levels.

[e] FACIT-F: Functional Assessment of Chronic Illness Therapy-Fatigue.

[f] HAQ: Health Assessment Questionnaire.

[g] EMS: Elderly Mobility Scale.

[h] PREDIMED: Prevención con Dieta Mediterránea.

[i] MD: Mediterranean diet.

Discussion

Principal Findings

To the best of our knowledge, this is the first study to test an eHealth intervention program for older adults who have food insecurity, which is a vulnerable population that is historically not exposed to health innovation. Before implementing a full-scale trial, we conducted this pilot study to explore the feasibility and acceptability of a multidisciplinary 12-week home-based intervention program focusing on improving dietary and physical activity through an interactive TV app. The results of this pilot study reveal aspects of the intervention that may need modification before moving to a full-scale trial to evaluate its effectiveness. This methodology is suggested whenever complex health interventions are developed and implemented in real-life settings [27]. Overall, the intervention program is feasible and highly acceptable. Our findings also provide insights into the promising effects of eHealth interventions on food insecurity, fatigue, and physical function.

During the recruitment period, 177 older adults meeting the eligibility criteria were identified; however, only 17.5% (31/177) were enrolled in the intervention program. This adherence rate was lower than initially anticipated [26] and can be explained by different reasons. First, our recruitment was restricted to selected primary health care centers during a specific period. Thus, we may have identified only those individuals for whom access to health care was not a problem, which may explain their low interest and motivation to engage in an eHealth intervention program at baseline. In addition, a recent study in a Portuguese context reported that older adults with food insecurity admit to reducing their medical visits and stopping medication for economic reasons [9]. Thus, a potential lack of awareness and low prioritization of managing food insecurity may reflect older adults' need to prioritize other activities perceived as more important. Thus, it might be important to broaden our recruitment strategy in future trials [46-49]. One excellent indicator for a future study, however, was that no participants were lost to follow-up during the study period. This retention rate contrasts with the dropout rate of 45% at the 3-month follow-up found by Van Doorn-Van Atten et al [47] in another study on the feasibility of an eHealth intervention targeting the nutritional status of vulnerable elderly people. Overall, our results suggest that when elderly individuals agree to participate, their engagement with the study will not be a barrier to trial completion.

A similar reason could explain our results regarding one of the primary study outcomes: the feasibility of the intervention program. Our monitoring protocol showed that all participants used the TV app; however, 32% (10/31) self-reported low use at the 3-month follow-up. Knowing that participant compliance is a common barrier to eHealth interventions [16], we integrated several strategies to monitor use and help participants remember to use the TV app (ie, monitoring the number of times participants accessed the TV app, technological reminders, phone calls, and questionnaires). However, these strategies are insufficient, which may present a challenge for future studies. To improve the monitoring process and thus increase participants' compliance, an emerging strategy is the use of pedometers or accelerometers, which have the potential to objectively and continuously monitor patients and inform researchers about participants' compliance with physical activity or exercise programs [20,21]. However, the adoption of this strategy in the present intervention program could complicate an already complex intervention while considering the characteristics of the end users, without guaranteeing that participants will use the monitoring instruments during their exercises or daily routine activities.

The baseline lack of participants' interest in using the TV app may be another explanation for self-reported low use. Even those participants who self-reported consistent use of the TV app stated that they had low interest in eHealth interventions in the baseline assessment. Nonetheless, after the intervention, we observed a substantial increase in participants' interest in eHealth interventions, both for improving food insecurity and for other purposes. This important result clearly justifies future investment in health innovation for this population.

In line with our feasibility results, the intervention program was highly accepted by participants both before and after the intervention. Acceptability, a multifaceted construct that reflects the extent to which people receiving a health care intervention consider it to be appropriate based on anticipated or experienced cognitive and emotional responses, was evaluated using constructs suggested by the theoretical framework of acceptability [37]. As expected, some of these constructs showed a significant decrease between the baseline and 3-month follow-up. Our eHealth intervention comprises a highly innovative digital program directed toward a vulnerable older adult population with low educational and socioeconomic status, low digital literacy, and a lack of regular exposure to health innovation, which are well-known barriers to eHealth adoption and implementation [24]. Thus, we anticipated that an innovative digital intervention would initially be met with high levels of acceptability. Later, however, as the intervention program involved active participation and commitment to physical exercise and a healthy diet, a decline in acceptability was observed. Despite this decrease, the median values of all acceptability constructs remained very high after the intervention.

Comparison of the present results with those of previous studies on the feasibility and acceptability of eHealth interventions focused on nutrition and physical activity among vulnerable older adults is challenging. The small number of previous studies and the high degree of heterogeneity in their designs, samples, interventions, and methods makes any comparison complex. Even so, our feasibility and acceptability results appear to be superior to those reported by Kraft et al [49] and Van Doorn-Van Atten et al [47].

Although impact analysis was not the primary objective of this study, the observed changes in key clinical outcomes revealed very promising short-term effects that support future randomized controlled trials. To the best of our knowledge, this is the first behavior change eHealth intervention that appears to effectively reduce food insecurity status and severity in a sample of older adults. The intervention program also showed improved fatigue and physical function of the participants, including their functional status, performance and mobility, and regularity of exercise. These results are in line with previous systematic reviews showing the effectiveness of eHealth interventions in improving the physical activity and physical functioning of older populations [20,21]. However, despite the systematic review by Marx et al [19] showing that eHealth malnutrition–related interventions improve the HRQoL and nutritional status of older adults, our intervention had no effect on these outcomes. These results could be explained by factors such as older age and greater physical, social, and economic vulnerabilities of our sample combined with the short follow-up period and use of a single instrument to evaluate the nutritional status. Longer follow-up periods and additional instruments for evaluating nutritional status (eg, 24-hour dietary recalls) should be considered for future effectiveness analysis.

Strengths and Limitations

This study has several limitations that need to be addressed in future full-scale trials. First, the low adherence of older adults may reflect selection bias. In addition to the large number of individuals who did not agree to participate in the study, 21 individuals were excluded because of technological problems with the installation of the TV app, which is an important limitation that needs to be addressed in future studies. Second, our feasibility and acceptability analyses were based exclusively on quantitative data collected through a structured questionnaire. The collection of qualitative data using semistructured interviews or focus groups could provide more detailed insights into participants' perspectives and thus reveal additional important aspects of the intervention in need of modification. Third, our short follow-up period prevented the analysis of the long-term effects of the intervention program. Fourth, the absence of a control group and the nonrandomized design of our study did not allow us to analyze the causal effects of the intervention program or to explore important feasibility aspects in preparation for a future randomized clinical trial, such as the randomization process and blinded outcome assessment. Finally, our pilot study involved a high proportion of missing data for almost all outcomes, which may compromise some of the data analysis performed.

Despite these limitations, this study's results can inform potential strategies for mitigating a growing global public health problem, which is expected to reach historic levels because of the COVID-19 pandemic [5,6]. Moreover, this pandemic is expected to cause an unprecedented social and economic crisis with devastating consequences for the most vulnerable populations [5,6]. Thus, there is an urgent need to develop feasible, acceptable, and effective interventions to promote healthy lifestyles and enhance well-being and quality of life among older adults. Our study represents the first step toward designing an effective intervention. Its strengths include the real-life health care context from which participants were recruited and the use of a validated instrument from the USDA Household Food Security Survey Module to identify eligible participants [29], which increases the external validity of our study. In addition, the study design was based on international frameworks for the design and implementation of complex and innovative interventions [27,28] and the assessment of acceptability constructs [37]. The strengths of this eHealth intervention program include (1) the early involvement of a multidisciplinary team and TV app end users in the development of the intervention program; (2) a multicomponent program aimed at improving several health outcomes (ie, nutritional, physical, and quality of life) based on behavioral change techniques; (3) use of a familiar ICT in participants' daily lives; and (4) close monitoring by the research team during the intervention period.

Conclusions

In conclusion, the findings of this pilot study reveal that our multidisciplinary 12-week home-based eHealth intervention program is a feasible and highly accepted method for improving the dietary and physical activity of older adults using a TV app, thus supporting a future full-scale trial. This intervention program not only reduced the proportion of older adults with food insecurity but also improved their fatigue and physical function.

Acknowledgments

This study was financed by national funds through the Foundation for Science and Technology, within the scope of the project DSAIPA/AI/0106/2019.

Conflicts of Interest

None declared.

References

1. Food and Agriculture Organization of the United Nations. The State of Food Security and Nutrition in the World 2019: Safeguarding against economic slowdowns and downturns. In: The State of Food Security and Nutrition in the World. New York, United States: United Nations; 2019:1-237 URL: http://www.fao.org/3/ca5162en/ca5162en.pdf
2. Coleman-Jensen A, Rabbitt MP, Smith MD, Gregory C, Singh A. Household food security in the United States in 2019. ERR-275, U.S. Department of Agriculture, Economic Research Service 2020 Sep:1-39 [FREE Full text]
3. Bickel G, Nord M, Price C, Hamilton W, Cook J. Guide to Measuring Household Food Security Revised 2000. Measuring Food Security in the United States. 2000. URL: https://fns-prod.azureedge.net/sites/default/files/FSGuide.pdf [accessed 2021-08-04]
4. Smith MD, Rabbitt MP, Coleman- Jensen A. Who are the world's food insecure? New evidence from the Food and Agriculture Organization's food insecurity experience scale. World Development 2017 May;93:402-412. [doi: 10.1016/j.worlddev.2017.01.006]
5. Gundersen C, Hake M, Dewey A, Engelhard E. Food insecurity during COVID-19. Appl Econ Perspect Policy 2020 Oct 02:13100 [FREE Full text] [doi: 10.1002/aepp.13100] [Medline: 33042509]
6. Wolfson JA, Leung CW. Food insecurity during COVID-19: an acute crisis with long-term health implications. Am J Public Health 2020 Dec;110(12):1763-1765. [doi: 10.2105/AJPH.2020.305953] [Medline: 32970451]
7. Fitzpatrick K, Harris C, Drawve G. Assessing U.S. Food Insecurity in the United States During COVID-19 Pandemic. 2020. URL: https://fulbright.uark.edu/departments/sociology/research-centers/community-family-institute/_resources/community-and-family-institute/revised-assessing-food-insecurity-brief.pdf [accessed 2021-08-04]
8. Gregório MJ, Rodrigues AM, Graça P, de Sousa RD, Dias SS, Branco JC, et al. Food insecurity is associated with low adherence to the Mediterranean diet and adverse health conditions in Portuguese adults. Front Public Health 2018;6:38 [FREE Full text] [doi: 10.3389/fpubh.2018.00038] [Medline: 29515992]

9. Fernandes SG, Rodrigues AM, Nunes C, Santos O, Gregório MJ, de Sousa RD, et al. Food insecurity in older adults: Results from the Epidemiology of Chronic Diseases Cohort Study 3. Front Med (Lausanne) 2018;5:203 [FREE Full text] [doi: 10.3389/fmed.2018.00203] [Medline: 30050904]
10. Hashemi R, Motlagh AD, Heshmat R, Esmaillzadeh A, Payab M, Yousefinia M, et al. Diet and its relationship to sarcopenia in community dwelling Iranian elderly: a cross sectional study. Nutrition 2015 Jan;31(1):97-104. [doi: 10.1016/j.nut.2014.05.003] [Medline: 25441592]
11. Arenas DJ, Thomas A, Wang J, DeLisser HM. A systematic review and meta-analysis of depression, anxiety, and sleep disorders in US adults with food insecurity. J Gen Intern Med 2019 Dec;34(12):2874-2882 [FREE Full text] [doi: 10.1007/s11606-019-05202-4] [Medline: 31385212]
12. Russell J, Flood V, Yeatman H, Mitchell P. Prevalence and risk factors of food insecurity among a cohort of older Australians. J Nutr Health Aging 2014 Jan;18(1):3-8. [doi: 10.1007/s12603-013-0339-6] [Medline: 24402381]
13. Portela-Parra ET, Leung CW. Food insecurity is associated with lower cognitive functioning in a national sample of older adults. J Nutr 2019 Oct 01;149(10):1812-1817 [FREE Full text] [doi: 10.1093/jn/nxz120] [Medline: 31240308]
14. Kihlström L, Burris M, Dobbins J, McGrath E, Renda A, Cordier T, et al. Food insecurity and health-related quality of life: a cross-sectional analysis of older adults in Florida, U.S. Ecol Food Nutr 2019;58(1):45-65. [doi: 10.1080/03670244.2018.1559160] [Medline: 30582362]
15. Global Observatory for eHealth. World Health Organization. 2018. URL: https://www.who.int/observatories/global-observatory-for-ehealth [accessed 2021-08-04]
16. van Den Berg N, Schumann M, Kraft K, Hoffmann W. Telemedicine and telecare for older patients— a systematic review. Maturitas 2012 Oct;73(2):94-114. [doi: 10.1016/j.maturitas.2012.06.010] [Medline: 22809497]
17. Eysenbach G. What is e-health? J Med Internet Res 2001 Jun;3(2):E20. [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
18. Takemoto M, Manini TM, Rosenberg DE, Lazar A, Zlatar ZZ, Das SK, et al. Diet and activity assessments and interventions using technology in older adults. Am J Prev Med 2018 Oct;55(4):105-115 [FREE Full text] [doi: 10.1016/j.amepre.2018.06.005] [Medline: 30241621]
19. Marx W, Kelly JT, Crichton M, Craven D, Collins J, Mackay H, et al. Is telehealth effective in managing malnutrition in community-dwelling older adults? A systematic review and meta-analysis. Maturitas 2018 May;111:31-46. [doi: 10.1016/j.maturitas.2018.02.012] [Medline: 29673830]
20. Muellmann S, Forberger S, Möllers T, Bröring E, Zeeb H, Pischke CR. Effectiveness of eHealth interventions for the promotion of physical activity in older adults: a systematic review. Prev Med 2018 Mar;108:93-110. [doi: 10.1016/j.ypmed.2017.12.026] [Medline: 29289643]
21. Stockwell S, Schofield P, Fisher A, Firth J, Jackson SE, Stubbs B, et al. Digital behavior change interventions to promote physical activity and/or reduce sedentary behavior in older adults: a systematic review and meta-analysis. Exp Gerontol 2019;120:68-87. [doi: 10.1016/j.exger.2019.02.020] [Medline: 30836130]
22. Yadav L, Haldar A, Jasper U, Taylor A, Visvanathan R, Chehade M, et al. Utilising digital health technology to support patient-healthcare provider communication in fragility fracture recovery: systematic review and meta-analysis. Int J Environ Res Public Health 2019 Oct 22;16(20):1-22 [FREE Full text] [doi: 10.3390/ijerph16204047] [Medline: 31652597]
23. Inglis SC, Clark RA, Dierckx R, Prieto-Merino D, Cleland JG. Structured telephone support or non-invasive telemonitoring for patients with heart failure. Cochrane Database Syst Rev 2015(10):CD007228. [doi: 10.1002/14651858.CD007228.pub3] [Medline: 26517969]
24. Hage E, Roo JP, van Offenbeek MA, Boonstra A. Implementation factors and their effect on e-Health service adoption in rural communities: a systematic literature review. BMC Health Serv Res 2013;13:19 [FREE Full text] [doi: 10.1186/1472-6963-13-19] [Medline: 23311452]
25. Depp CA, Schkade DA, Thompson WK, Jeste DV. Age, affective experience, and television use. Am J Prev Med 2010 Aug;39(2):173-178 [FREE Full text] [doi: 10.1016/j.amepre.2010.03.020] [Medline: 20621265]
26. Rodrigues AM, Gregório MJ, Gein P, Eusébio M, Santos MJ, de Sousa RD, et al. Home-based intervention program to reduce food insecurity in elderly populations using a TV app: study protocol of the randomized controlled trial Saúde.come senior. JMIR Res Protoc 2017 Mar 13;6(3):e40 [FREE Full text] [doi: 10.2196/resprot.6626] [Medline: 28288956]
27. 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]
28. Stilgoe J, Owen R, Macnaghten P. Developing a framework for responsible innovation. Res Policy 2013 Nov;42(9):1568-1580. [doi: 10.1016/j.respol.2013.05.008]
29. Gregório M, Graça P, Nogueira P, Gomes S, Santos C, Boavida J. Proposta Metodológica para a Avaliação da Insegurança Alimentar em Portugal. Rev Nutrícias. 2014. URL: https://repositorio.ul.pt/handle/10451/17617 [accessed 2021-08-04]
30. Prochaska JO, DiClemente CC, Norcross JC. In search of how people change. Applications to addictive behaviors. Am Psychol 1992 Sep;47(9):1102-1114. [doi: 10.1037//0003-066x.47.9.1102] [Medline: 1329589]
31. Gregório MJ, Rodrigues AM, Eusébio M, Sousa RD, Dias S, André B, et al. Dietary patterns characterized by high meat consumption are associated with other unhealthy life styles and depression symptoms. Front Nutr 2017;4:25 [FREE Full text] [doi: 10.3389/fnut.2017.00025] [Medline: 28660194]

32. Portuguese Food Balance - 2012 - 2016. Instituto Nacional de Estatística. 2017. URL: https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_publicacoes&PUBLICACOESpub_boui=289818234&PUBLICACOESmodo=2 [accessed 2021-08-04]
33. Santos DM, Rodrigues SS, de Oliveira BM, Vaz de Almeida MD. Diet quality in elderly Portuguese households. J Nutr Health Aging 2014 Mar;18(3):243-250. [doi: 10.1007/s12603-013-0431-y] [Medline: 24626750]
34. A Saúde dos Portugueses. Perspetiva 2015. Direção-Geral da Saúde. 2015. URL: https://www.omd.pt/content/uploads/2017/12/dgs-saude-portugueses-2015.pdf [accessed 2021-08-04]
35. Trust for America's Health. Recommendations to Promote Health and Well-being Among Aging Populations. Washington, D.C: Altarum Institute; 2012.
36. Pais-Ribeiro J, Silva I, Ferreira T, Martins A, Meneses R, Baltar M. Validation study of a Portuguese version of the Hospital Anxiety and Depression Scale. Psychol Health Med 2007 Mar;12(2):225-235. [doi: 10.1080/13548500500524088] [Medline: 17365902]
37. Sekhon M, Cartwright M, Francis JJ. Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. BMC Health Serv Res 2017 Jan 26;17(1):88 [FREE Full text] [doi: 10.1186/s12913-017-2031-8] [Medline: 28126032]
38. Ferreira LN, Ferreira PL, Pereira LN, Oppe M. EQ-5D Portuguese population norms. Qual Life Res 2014 Mar 3;23(2):425-430. [doi: 10.1007/s11136-013-0488-4] [Medline: 23912856]
39. Ferreira LN, Ferreira PL, Pereira LN, Oppe M. The valuation of the EQ-5D in Portugal. Qual Life Res 2014 Mar 8;23(2):413-423. [doi: 10.1007/s11136-013-0448-z] [Medline: 23748906]
40. Chandran V, Bhella S, Schentag C, Gladman DD. Functional assessment of chronic illness therapy-fatigue scale is valid in patients with psoriatic arthritis. Ann Rheum Dis 2007 Jul;66(7):936-939 [FREE Full text] [doi: 10.1136/ard.2006.065763] [Medline: 17324972]
41. Santos R, Reis P, Rebelo L, Dias F, Rosa C, Queiroz M. "Health Assessment Questionnaire" (versão curta): adaptação para língua portuguesa e estudo da sua aplicabilidade. Acta Reuma Port 1996;76:21-26 [FREE Full text]
42. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. Arthritis Rheum 1980 Feb;23(2):137-145. [doi: 10.1002/art.1780230202] [Medline: 7362664]
43. Santos A, Ramos N, Estêvão P, Lopes A, Pascoalinho J. Instrumentos de Medida Úteis no Contexto da Avaliação em Fisioterapia. Re(habilitar)-Revista da ESSA. 2005. URL: https://scholar.google.com/citations?view_op=view_citation&hl=pt-BR&user=MJN-oH0AAAAJ&citation_for_view=MJN-oH0AAAAJ:isU91gLudPYC [accessed 2021-08-04]
44. Yu MS, Chan CC, Tsim RK. Usefulness of the Elderly Mobility Scale for classifying residential placements. Clin Rehabil 2007 Dec;21(12):1114-1120. [doi: 10.1177/0269215507080789] [Medline: 18042607]
45. Schröder H, Fitó M, Estruch R, Martínez-González MA, Corella D, Salas-Salvadó J, et al. A short screener is valid for assessing Mediterranean diet adherence among older Spanish men and women. J Nutr 2011 Jun;141(6):1140-1145 [FREE Full text] [doi: 10.3945/jn.110.135566] [Medline: 21508208]
46. van Doorn-van Atten MN, de Groot LC, Romea AC, Schwartz S, de Vries JH, Haveman-Nies A. Implementation of a multicomponent telemonitoring intervention to improve nutritional status of community-dwelling older adults: a process evaluation. Public Health Nutr 2019 Feb;22(2):363-374. [doi: 10.1017/S1368980018002185] [Medline: 30175698]
47. van Doorn-van Atten MN, Haveman-Nies A, Heery D, de Vries JH, de Groot LC. Feasibility and effectiveness of nutritional telemonitoring for home care clients: a pilot study. Gerontologist 2019 Jan 09;59(1):158-166. [doi: 10.1093/geront/gny059] [Medline: 29860379]
48. Hirani SP, Beynon M, Cartwright M, Rixon L, Doll H, Henderson C, et al. The effect of telecare on the quality of life and psychological well-being of elderly recipients of social care over a 12-month period: the Whole Systems Demonstrator cluster randomised trial. Age Ageing 2014 May;43(3):334-341. [doi: 10.1093/ageing/aft185] [Medline: 24333802]
49. Kraft M, van den Berg N, Kraft K, Schmekel S, Gärtner S, Krüger J, et al. Development of a telemedical monitoring concept for the care of malnourished geriatric home-dwelling patients: a pilot study. Maturitas 2012 Jun;72(2):126-131. [doi: 10.1016/j.maturitas.2012.02.011] [Medline: 22440535]

Abbreviations

EMS: Elderly Mobility Scale
EQ-5D-3L: European Quality of Life Questionnaire with five dimensions and three levels
FACIT-F: Functional Assessment of Chronic Illness Therapy-Fatigue
HAQ: Health Assessment Questionnaire
HRQoL: health-related quality of life
ICT: information and communication technology
MD: Mediterranean diet
PREDIMED: Prevención con Dieta Mediterránea
TV: television

USDA: United States Department of Agriculture

Edited by R Kukafka; submitted 31.12.20; peer-reviewed by N Vandenberg, L Yadav; comments to author 15.02.21; revised version received 19.03.21; accepted 06.05.21; published 31.08.21.

Please cite as:
Gomes LA, Gregório MJ, Iakovleva TA, Sousa RDD, Bessant J, Oliveira P, Branco JC, Canhão H, Rodrigues AM
A Home-Based eHealth Intervention for an Older Adult Population With Food Insecurity: Feasibility and Acceptability Study
J Med Internet Res 2021;23(8):e26871
URL: https://www.jmir.org/2021/8/e26871
doi:10.2196/26871
PMID:34463638

©Luís Antunes Gomes, Maria João Gregório, Tatiana A Iakovleva, Rute Dinis de Sousa, John Bessant, Pedro Oliveira, Jaime C Branco, Helena Canhão, Ana Maria Rodrigues. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 31.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Patient and Professional Experiences With Virtual Antenatal Clinics During the COVID-19 Pandemic in a UK Tertiary Obstetric Hospital: Questionnaire Study

Lauren Marie Quinn[1], MBChB, BMedSci; Oluwafumbi Olajide[2], MBBS; Marsha Green[2], MBBS; Hazem Sayed[2], MBBS; Humera Ansar[2], MBBS

[1]University Hospitals of Leicester, Leicester, United Kingdom
[2]Department of Obstetrics, University Hospitals of Leicester, Leicester, United Kingdom

Corresponding Author:
Lauren Marie Quinn, MBChB, BMedSci
University Hospitals of Leicester
Infirmary Square
Leicester, LE1 5WW
United Kingdom
Phone: 44 03003031573
Email: lauren.quinn6@nhs.net

Abstract

Background: The COVID-19 pandemic required rapid implementation of virtual antenatal care to keep pregnant women safe. This transition from face-to-face usual care had to be embraced by patients and professionals alike.

Objective: We evaluated patients' and professionals' experiences with virtual antenatal clinic appointments during the COVID-19 pandemic to determine satisfaction and inquire into the safety and quality of care received.

Methods: A total of 148 women who attended a virtual antenatal clinic appointment at our UK tertiary obstetric care center over a 2-week period provided feedback (n=92, 62% response rate). A further 37 health care professionals (HCPs) delivering care in the virtual antenatal clinics participated in another questionnaire study (37/45, 82% response rate).

Results: We showed that women were highly satisfied with the virtual clinics, with 86% (127/148) rating their experience as good or very good, and this was not associated with any statistically significant differences in age (P=.23), ethnicity (P=.95), number of previous births (P=.65), or pregnancy losses (P=.94). Even though 56% (83/148) preferred face-to-face appointments, 44% (65/148) either expressed no preference or preferred virtual, and these preferences were not associated with significant differences in patient demographics. For HCPs, 67% (18/27) rated their experience of virtual clinics as good or very good, 78% (21/27) described their experience as the same or better than face-to-face clinics, 15% (4/27) preferred virtual clinics, and 44% (12/27) had no preference. Importantly, 67% (18/27) found it easy or very easy to adapt to virtual clinics. Over 90% of HCPs agreed virtual clinics should be implemented long-term.

Conclusions: Our study demonstrates high satisfaction with telephone antenatal clinics during the pandemic, which supports the transition toward widespread digitalization of antenatal care suited to 21st-century patients and professionals.

(J Med Internet Res 2021;23(8):e25549) doi:10.2196/25549

KEYWORDS

antenatal; virtual clinic; technology; COVID-19; United Kingdom; pandemic; feasibility; effective; telehealth; virtual health

Introduction

The COVID-19 pandemic presented challenges to obstetric departments worldwide, resulting in increased pressures on the delivery of routine antenatal care. Consensus guidelines recommended that pregnant women should self-isolate and avoid coming into the hospital unless necessary to avoid contracting COVID-19 [1-3]. Following the pandemic, the Royal College of Obstetricians and Gynaecologists recommended a minimum of six antenatal appointments (a reduction from the usual eight face-to-face visits). However, this only accounted for low-risk pregnancies, as higher risk women still required specialist antenatal clinic appointments. Royal College of Obstetricians and Gynaecologists therefore also advised that telephone or video consultations were safe, and their adoption should be maximized to minimize unnecessary contact between patients and care providers [4,5].

Face-to-face antenatal care has long been recognized as the standard of care because of the importance of routine screening for blood pressure and urine assessment, and to provide personalized support. However, telemedicine approaches in perinatal care have been implemented with success [6], and the National Health Service (NHS) England's long-term plan is a digital agenda aiming to reduce face-to-face appointments by a third over the next 5 years [7]. Pflugieson and Mou [8] compared virtual and traditional antenatal appointments and showed that patient satisfaction was significantly higher in women who received virtual care compared to face-to-face appointments, and there was no difference in preference between telephone or video calls for the virtual clinics. However, during the COVID-19 pandemic, when the majority of antenatal care was rapidly transformed to virtual care in the United Kingdom, it was unclear how pregnant women felt about this shift in culture.

In addition to the patient experience of virtual clinics, it is also important to consider the clinicians' perspective. A mixed-methods acceptability study evaluating virtual clinics at the micro, meso, and macro levels found that technical challenges could be prohibitive to staff running virtual clinics [9]. However, when clinical, technical, and practical requirements were achieved, staff were satisfied and felt virtual clinics were safe [9]. Nevertheless, in the setting of a global pandemic, virtual clinics had to become the new normality for antenatal care, providing the advantage of instilling digitalized antenatal care into obstetric care systems worldwide [10].

The aim of this study is to evaluate patient and health care professional (HCP) satisfaction, preferences, and experiences of a virtual antenatal clinic during the COVID-19 pandemic from a tertiary obstetric hospital in the United Kingdom. This was important to understand more representative viewpoints from pregnant women when virtual care was the default model of care due to the pandemic, as opposed to virtual care being made available as an option instead of or in addition to traditional face-to-face care as previous studies have explored.

Methods

Study Design

In March 2020, antenatal face-to-face general and subspecialist clinic appointments in our regional tertiary obstetric unit were transformed to virtual telephone clinic appointments. Telephone consultations were conducted by either consultants, registrars and junior obstetric doctors, or midwives. The junior obstetric clinic team members discussed all cases with the consultant lead during or at the end of the virtual clinic, and treatment decisions were signed off by the consultant, for example, confirmed mode of delivery and intrapartum care plans (this meant some women may have received more than one call to confirm management plan). The process of setting up the virtual antenatal clinic at our center is reported elsewhere [11].

Patient Experience

An anonymized questionnaire was used to evaluate satisfaction with the virtual clinic experience. The questionnaire was adapted, with consent, from the questionnaire validated by Pflugieson and Mou [8] (Multimedia Appendix 1). Women rated each question on a Likert scale, from very poor to very good, in relation to scheduling, technology, HCP rating, patient-orientated nature, overall rating, and preferences. There were 16 Likert scale questions. Study participants were finally asked to outline the benefits of the virtual clinic and areas of improvement by free text. The following demographic variables were collected: age, ethnicity, number of previous births, and number of pregnancy losses (see Multimedia Appendix 1 for the patient questionnaire). In light of the pandemic, blood pressure and doppler assessments were organized and performed by the community midwives, as it was not deemed possible to train or deliver this equipment to patients safely.

Women were verbally consented to participate in the questionnaire study during the telephone consultation and could provide feedback by email or telephone questionnaire within 2 weeks of their clinic appointment. A pilot study (n=4 women) was performed to test the questionnaire and acceptability of the study design, and the questions were modified as a result, but this data was not included for analysis.

The questionnaires were collected from women who had a virtual clinic consultation over a 2-week period between Monday, May 4 and Friday, May 15, 2020. This period was chosen because the virtual clinics had been running and optimized over a month prior to this. All questionnaire responses were anonymized, and the feedback was collected independently from the clinic staff. For non-English speaking women, a staff member who could speak the required language (Gujurati, Hindi, or Urdu) obtained feedback by translating over the telephone.

Health Care Professional Experience

To understand HCPs' experiences of virtual antenatal clinics, two surveys were produced: the first aimed at clinical staff conducting the virtual antenatal clinics (doctors and midwives: HCP survey) and the second aimed at clerical staff organizing the virtual clinics (administrator survey). The HCP survey included 46 questions, ranked on a Likert scale regarding how safe, effective, efficient, and satisfied HCPs were with the virtual clinics and how this compared to their experience of face-to-face clinics. There were also questions enquiring into the benefits and consequences of virtual clinics and their opinions on how well the virtual clinics had been implemented at this center. Finally, HCPs were asked to give their preference for virtual or face-to-face appointments.

The administrator survey consisted of 26 questions exploring experiences of coordinating the virtual clinics, how this compared to face-to-face clinics, and whether they felt the virtual clinic system was effective.

HCPs were surveyed over a 2-week period in July 2020 and were asked to comment on their experiences from the preceding 3 months. This later time period was chosen because from April to mid-June the obstetric clinic team consisted of 15 individuals who were nonpatient facing but who could undertake the clinics full time. However, from June onwards, the frontline clinical staff started to come back to work in the virtual clinic environment. Therefore, to increase the sample size, we delayed the HCP and administrator surveys to July, when more staff had experienced the new virtual clinic environment, which they could compare to the traditional face-to-face clinics. Importantly, the running of the virtual clinics did not change between the patient survey and the professional survey.

The two surveys were tested on a pilot population of 4 individuals, and the questionnaires were adapted accordingly. A list of all the HCPs and administrators who had undertaken virtual obstetric clinics in the preceding 12 weeks was compiled by the study team (n=45). The questionnaire was sent out via email using Survey Monkey. All questionnaire responses were anonymous (see Multimedia Appendix 2 for the HCP questionnaire).

Ethics

This study was registered as a service evaluation and quality improvement project with the local audit department (reference number: 10560a).

Statistical Analysis

Descriptive statistics were performed in Excel (Microsoft Corporation). The Mann-Whitney *U* and chi-square tests were used to compare demographics between responses to the virtual clinic experience. Statistical analysis was performed in Graphpad Prism (GraphPad Software, Inc) and SPSS v22 (IBM Corp). The level of statistical significance was set at $P<.05$.

Results

Patients' Experience of Virtual Antenatal Clinics

In the 2-week period evaluated, 268 women had a virtual consultation and a further 45 women did not attend or cancelled their virtual clinic appointment. Of the 268 women, 28 did not consent to participate in the study. Of the 240 women who were seen in the virtual clinic and consented to providing feedback, 148 completed and returned the questionnaire, resulting in a 62% response rate (see flowchart in Multimedia Appendix 3).

Demographics

Of those women who completed the questionnaire, the mean age was 31 (SD 5.829) years. The majority of women were Caucasian (168/148, 78.5%), 20% (30/148) were South Asian, and 1.5% (2/148) were another ethnic group. The majority of women were multipara (132/148, 89.5%), and 10.5% (16/148) of women were primipara. In addition, the majority of women had no previous pregnancy loss (99/148, 67%), and 33% (50/148) had one or more self-reported previous pregnancy losses (miscarriage, ectopic, or molar pregnancy).

Quantitative Analysis: Descriptive

Satisfaction with the virtual clinic appointment was rated as good or very good by 86% (127/148) of women (Figure 1), and 82% (112/148) were very likely or likely to recommend a virtual clinic appointment based on their experience (Figure 1). In addition, 96% (142/148) of women rated the overall quality of the virtual clinic appointment, excluding the technology, component as 6 out of 10 or higher (Figure 2).

Figure 1. Patient satisfaction and likelihood of recommending virtual antenatal clinics. UHL: University Hospitals of Leicester.

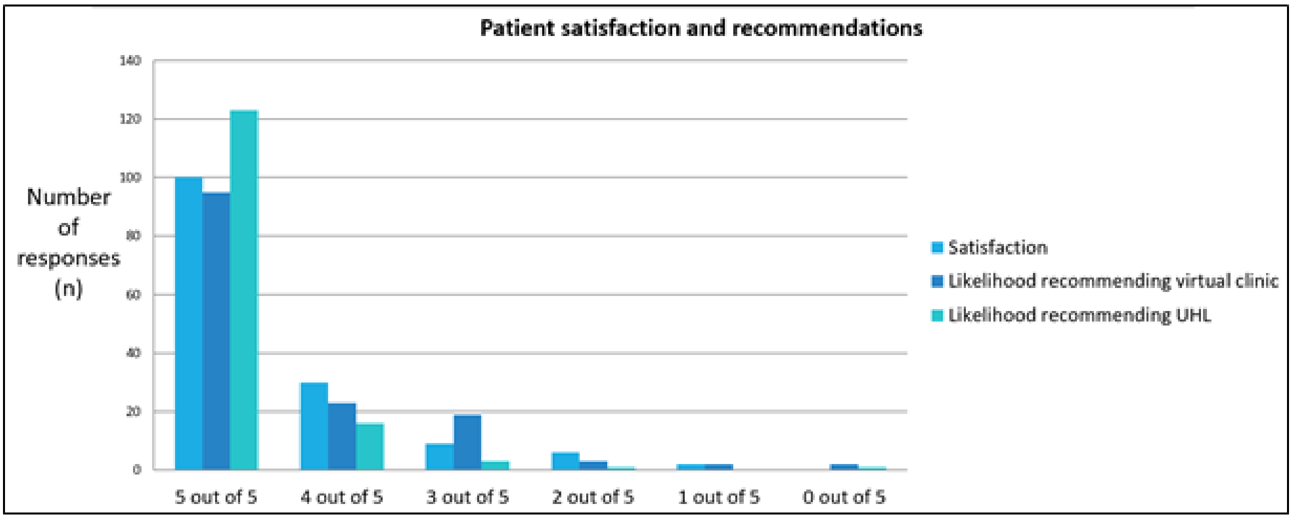


Figure 2. Patient denoted overall quality rating for the virtual antenatal clinics, including and excluding technology elements.

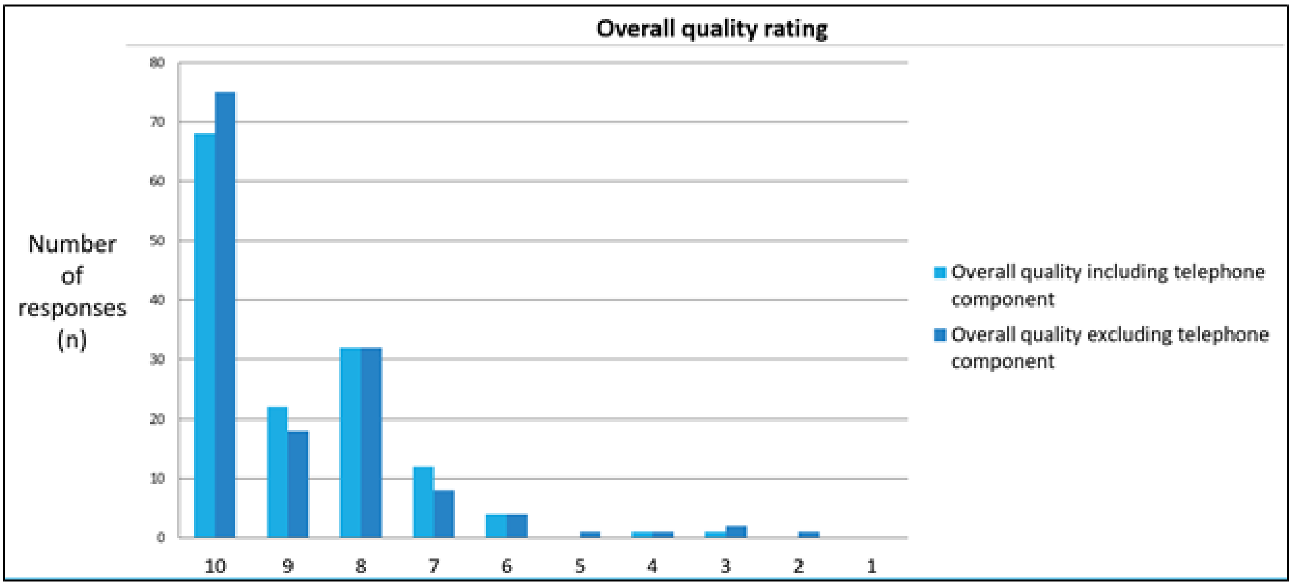


Table 1 shows the questions asked in the study questionnaire about their experience of the virtual clinic (adapted from Pflugieson and Mou [8]) and the mean score out of five with the SD provided.

Of those surveyed, 38% (56/148) remembered being asked COVID-19 screening questions, and 12% (18/148) could not remember being asked, leaving 50% (74/148) who were not asked COVID-19 screening questions during the virtual clinic appointment. Almost all women (144/148, 97%) said the consultation felt private to them. When participants were asked to select between virtual and face-to-face appointments, 25% (37/148) preferred virtual, 10% (15/148) had no preference, and 9% (13/148) said it was dependent on other factors such as the ongoing pandemic, giving a total of 44% (65/148) of women who would be happy with virtual appointments. However, 56% (83/148) of respondents still preferred face-to-face appointments.

Table 1. Questions asked in the questionnaire with the mean and SD for the Likert score for each rating.

Question number	Question	Likert rating out of 5, mean (SD)
1	Ease of scheduling your virtual clinic appointment	4.5 (0.8)
2	Convenience of virtual clinic times and dates	4.6 (0.8)
3	Ease of connecting for your virtual appointments	4.8 (0.5)
4	Quality of connection during virtual appointments	4.8 (0.5)
5	How well the doctor explained her role in your care	4.4 (1.0)
6	Friendliness/courtesy of doctor	4.8 (0.7)
7	Explanation of plan for next appointment(s) and follow up	4.6 (0.7)
8	Skill and knowledge of the doctor	4.7 (0.7)
9	Degree to which the doctor took time to listen to you	4.6 (0.9)
10	Degree to which doctor helped you to make informed decisions	4.4 (1.0)
11	Doctor’s concern for and ability to answer your questions and worries	4.5 (1.0)
12	Satisfaction with virtual appointments	4.5 (0.9)
13	Likelihood of recommending virtual appointments/your prenatal care doctor	4.4 (1.0)
14	Likelihood that you will continue to seek care from the University Hospitals of Leicester	4.8 (0.6)
15	Overall quality, inclusive of the technology element on a scale of 1-10	8.9 (1.3)
16	Overall quality, exclusive of the technology element on a scale of 1-10	8.9 (1.5)

Statistical Analysis

There were no statistically significant differences between virtual clinic satisfaction rating (out of five) and age (*P*=.23), ethnicity (*P*=.95), parity (primipara or multiparous; *P*=.65), or number of previous pregnancy losses (none or ≥1; *P*=.94). There were also no statistically significant differences between preference for clinic type (ie, virtual or face-to-face clinic) with age (*P*=.07), ethnicity (*P*=.93), smoking status (*P*=.78), parity (primipara or multiparous; *P*=.79), or number of previous pregnancy losses (none or ≥1; *P*=.54). When ranking the overall quality of the virtual clinic including the technology components, primipara women were significantly more likely to rate it as 10 out of 10 compared to multiparous women (10/13, 78% vs 59/115, 51%; *P*=.03), whereas multiparous women were significantly more likely to rate it as 7 to 9 out of 10 compared to primiparous women (48/115, 42% vs 2/13, 15%; *P*=.03).

Summary of Free-Text Responses

The key benefits highlighted were convenience, avoiding travel, being able to stay at home, and staying safe. Patients also said that the communication from their doctor was good, and the doctors took time to listen to them, were friendly, and gave good explanations. A few women expressed initial concern about the concept of a virtual clinic but were happy with the experience once they had tried it.

In terms of areas for improvement, a lot of women said they would not change anything. However, the main improvement suggested was that the virtual clinic consultations should be at the specific time given. Some women felt they should have been given a choice whether they wanted a face-to-face or virtual appointment.

Of the minority of women who were dissatisfied with their virtual clinic consultation, rating the consultation as very poor or poor, the main issues raised were around the doctor not knowing them and their history well, and wanting a face-to-face appointment because of their individual circumstances, for example, to physically examine them. Timing of the clinic appointment was also a common issue, in terms of having their telephone consultation later than the expected time.

Health Care Professionals' Experience of Virtual Antenatal Clinics

Demographics

A total of 37 staff members completed the questionnaires, of which 27 completed the HCP survey and 10 completed the administrators survey; this was an 82% response rate overall of the 45 staff members who had performed virtual obstetric clinic duties in the time period evaluated.

Of the 27 HCPs who completed the HCP survey, 38% (n=10) were consultants, 38% (n=10) were registrars, 16% (n=4) were junior doctors, and 4% (n=1) were midwives. In terms of obstetric experience, 33% (n=9) had more than 21 years of experience, 19% (n=5) had 11 to 20 years of experience, and 26% (n=7) had less than 10 years of obstetric experience. Over half of the HCPs (n=15, 54%) had done more than 30 virtual consultations in the study period, 11% (n=3) had done 11 to 30 virtual consultations, and 23% (n=6) had done fewer than 10 consultations.

Quantitative Analysis: Descriptive

In terms of HCPs' experience of the virtual antenatal clinics, 67% (n=18) had a good or very good experience of virtual clinics and 78% (n=21) described their experience as the same or better than face-to-face clinics. Although 74% (n=20) of clinicians received no training, 67% (n=18) felt it was very easy or easy to adapt to the virtual clinics. Nonetheless, 56% felt training would have been helpful or very helpful. A total 78% (n=21) of participants felt the quality of connection was good or very good during the virtual clinics.

The majority felt virtual clinics were safe (n=22, 82%), 100% (n=27) felt virtual clinics were effective at delivering on high quality patient care, and 89% (n=24) perceived the care received by patients to be better or comparable to face-to-face appointments. Furthermore, 56% (n=15) felt it was easier or just as easy to seek advice or a second opinion in virtual clinics.

Although 74% (n=20) perceived virtual clinics took longer than face-to-face appointments, 63% (n=17) felt virtual clinics were as or more efficient than face-to-face clinics overall. Within each 4- to 4.5-hour clinic period, 70% (n=19) of HCPs could review on average 5 to 8 patients in a virtual clinic, while 22% (n=6) could consult with more than 9 patients. Importantly, 93% (n=25) felt virtual clinics should be implemented in the long-term. In terms of clinician preference, 44% (n=12) gave no preference, 15% (n=4) preferred virtual clinics, and 27% (n=7) preferred face-to-face appointments.

In terms of clerical staffs' experience of coordinating the virtual antenatal clinics (n=10), 70% (n=7) found it easy or very easy to schedule the virtual clinic appointments for patients, but 50% (n=5) felt the virtual clinics took more time to complete the clinic outcomes for patients. All (n=10, 100%) felt the care received by patients was the same or better with the virtual clinics compared to the face-to-face appointments, and 80% (n=8) felt virtual clinics were safe for patients. Of these, 80% (n=8) rated their experience of virtual clinics as good or very good, and 80% (n=8) said the virtual clinics in their current format would be feasible for the future and should be implemented long-term. Finally, 60% (n=6) had no preference for virtual or face-to-face clinics, and 90% (n=9) found it easy or very easy to adapt to the virtual clinics.

HCPs' perceived benefits of virtual antenatal clinics included patient convenience, environmentally friendly, and cost-effectiveness. Further benefits that ranked highly include the improved efficiency of virtual clinics, staff convenience, and patient-centered approach.

The principal barriers included the unavoidable need for face-to-face appointments in certain cases, limitations with technology, difficulties embedding the virtual clinics into the systems and process, and the adaptation required by clerical and clinical teams alike.

Summary of HCP Free-Text Responses

Key themes from the HCPs that arose were that the virtual clinics reduced unnecessary visits to the hospital for patients and allowed low-risk pregnancies to be managed safely from the patient's own home. Many felt patient compliance was better and perceived fewer *did not attend* appointments, and the clinics were considered to be patient-centered with good continuity of care. A concern raised was around a reduction in the training for junior doctors in virtual clinics compared to face-to-face clinics; however, others felt team cohesion was better, and they were able to discuss more cases with senior colleagues during the virtual clinics.

Discussion

Main Findings

We demonstrated that women were highly satisfied with all aspects of their virtual clinic experience and that this did not differ with age, ethnicity, previous pregnancies, or previous pregnancy losses. Although 56% (83/148) of women in our study would prefer a face-to-face clinic appointment, 44% (65/148) preferred virtual clinics or had no preference. We also showed that HCPs and administrative staff in our center were highly satisfied with virtual antenatal clinics and felt they were at least comparable to face-to-face appointments in terms of safety, care received, efficiency, and experience.

The COVID-19 pandemic has provided a much needed opportunity to digitalize antenatal care; the vast majority of pregnant women own a smartphone [12] and commonly seek knowledge from online sources [13]. eHealth has been shown to confer benefits to lifestyle and mental health outcomes for pregnant women, who report that eHealth interventions are convenient and acceptable [14]. Hence, there is scope to increasingly integrate virtual and telemedicine approaches to bring antenatal care into the 21st century [15].

Our finding that virtual antenatal clinics were acceptable to all patients surveyed regardless of sociodemographic differences is supported by Pflugieson and Mou [8] who showed similarly high satisfaction with virtual care. Historically, women's uptake for virtual care has been limited due to concerns about lack of perceived support and long gaps between in-person visits [16,17]. However, a shift in practice with the pandemic has allowed increased capture of a wider proportion of women's preferences. Holcomb et al [18] showed high patient satisfaction with audio-only virtual antenatal care during the COVID-19 pandemic, demonstrating that 99% of women felt their needs were met with virtual care, and compliance with virtual clinics (88%) was significantly higher than in-person appointments (82%; $P<.001$). A cross-sectional study by Futterman et al [19] compared virtual with in-person appointments and found high satisfaction with both, although in-person satisfaction was significantly higher. We therefore acknowledge, similar to Pflugieson and Mou [8], that women should ideally be offered a choice in their antenatal care modality because of the unique benefits received from patient-centered, face-to-face contact with a HCP. Aziz et al [6] similarly reported on the importance of combining face-to-face and telemedicine approaches for high-risk pregnancies during the pandemic, but we must ensure adoption of telemedicine strategies that do not compromise materno-fetal outcomes. A randomized controlled trial by Butler Tobah et al [20] compared alternative prenatal care (with fewer on-site visits and more virtual appointments) with usual face-to-face care and showed that women had higher levels of satisfaction and less stress with the virtual care arm, with no difference in materno-fetal outcomes or perceived quality of care. These findings have been further supported by systematic reviews and studies that have confirmed safety and efficacy of virtual clinics and telehealth despite a reduction of in-person visits [21-23]. Where face-to-face appointments cannot be avoided, Dashraath et al [24] outlined how in-person antenatal care can be safely practiced in the context of the COVID-19 pandemic, with social distancing and appropriate personal protective equipment for staff and women alike [24]. However, despite these important measures, Fryer et al [15] acknowledged the further work that needs to be done to protect higher risk pregnancies and ensure health inequality gaps are not widened.

The high professional satisfaction with virtual care reported in our study can be attributed to the simple model of telephone consultations that we adopted, which minimized the technical issues experienced. Greenhalgh et al [9] performed a multilevel analysis of virtual clinics, pre–COVID-19, in which they experienced significant technical issues that were prohibitive to implementation, resulting in a virtual consultation rate of only 22%. Nevertheless, our study is the first, to our knowledge, to evaluate the HCP perspective of virtual clinics during and before the COVID-19 pandemic. We found good integration between clinical and administrative teams to ensure the clinics ran efficiently and meant satisfaction from both teams was high. The finding that our virtual clinics were perceived to take longer than face-to-face appointments is not supported by other studies [9,16] and may be explained by inexperience or apprehension around needing to adapt to the new clinic experience, which is recognized as a significant challenge for HCPs; staff were asked to transition to a completely new way of working without training beforehand [9]. However, given that virtual clinics were considered to be more efficient than face-to-face appointments, this would suggest time was saved elsewhere. The majority of staff surveyed were in support of training for virtual clinics, which is not routinely part of the curriculum, and we anticipate this would further improve efficiency, satisfaction, and ease of adaptation to virtual clinics.

The benefits of virtual clinics are widely reported to include the environmentally friendly nature, patient-centered approach, and opportunity to build better rapport between patient and professionals. Furthermore, the challenges presented such as inability to examine can be overcome with video consultations [25]. Therefore, the benefits combined with the high satisfaction reported here demonstrate the importance of integrating virtual clinics into obstetric care services in the longer term to align with the NHS digital long-term plan [7].

Strengths and Limitations

We achieved a 62% (92/148) questionnaire response rate for women who received a virtual clinic consultation and 82% (37/45) response rate for the professionals surveyed in a large UK tertiary obstetric care center during the study period. Pflugieson and Mou [8] attained a 12.1% to 19.8% response rate in their study [8], hence reflecting the widely representative nature of our study findings.

Limitations of the study include its cross-sectional nature, as we only evaluated patient satisfaction over a 2-week period, and collection of the feedback was retrospective. There was also risk of selection bias in terms of the population of women who chose to complete the study questionnaire, but the majority response rate (92/148, 62%) will have negated this effect. The HCP sample size (n=37) was small because of the single center nature of this study, but attaining an 82% (37/45) response rate will have provided representative perspectives from this small cohort. Future work should be performed in larger, more diverse cohorts of HCPs and patients to further elucidate our study findings and to better differentiate virtual clinic preferences. We were unable to compare virtual clinic experiences directly with face-to-face care because of the pandemic, but as the lockdown restrictions ease, this is an area we will be evaluating moving forward to assess the long-term feasibility of virtual antenatal clinics. It is important to note that a minority of our cohort were primipara women, and the majority had no previous pregnancy loss, which reflects a lower risk population and may partly account for the high satisfaction with virtual compared to face-to-face consultations. Furthermore, this questionnaire was only designed to evaluate preferences, and further work needs to focus on safety, cost-effectiveness, and maternal and fetal outcomes of virtual care compared to face-to-face antenatal clinics. Finally, it must be acknowledged that satisfaction was not 100%, and for a minority of patients and professionals, virtual care and clinics were not acceptable. In these instances, a combination of virtual and face-to-face care would be a necessary approach, which aligns with the consensus viewpoint for antenatal care services.

Conclusion

We have shown that, despite rapid transformation and implementation of virtual antenatal clinics during the COVID-19 pandemic, patient and professional satisfaction with this service was very high. The virtual antenatal clinics have been widely accepted by women regardless of sociodemographic differences, which supports feasibility of the virtual clinics moving forward. Our study supports integration of virtual antenatal clinics alongside face-to-face delivery of care as and where appropriate to ensure delivery of patient-centered care. Further telemedicine strategies that aim to personalize care for pregnant women warrant further exploration.

Acknowledgments

A special thank you to Ms Penelope McParland, Ms Helena Maybury, Mr Eamonn Breslin (consultant obstetricians), and Ms Nicola Hooson (obstetric clinical secretary) for their invaluable contributions.

Authors' Contributions

LMQ, OO, HS, MG, and HA performed the data collection. LMQ, OO, and HA performed the data analysis. LMQ, OO, HS, MG, and HA wrote the paper, and all authors reviewed the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Virtual antenatal clinic survey for pregnant women.
[DOCX File , 15 KB - jmir_v23i8e25549_app1.docx]

Multimedia Appendix 2
Virtual antenatal clinic survey for professionals.
[DOCX File , 19 KB - jmir_v23i8e25549_app2.docx]

Multimedia Appendix 3
CONSORT flow diagram.
[PNG File , 62 KB - jmir_v23i8e25549_app3.png]

References

1. Coronavirus (COVID-19), pregnancy and women's health. Royal College of Obstetrics and Gynaecology. 2020. URL: https://www.rcog.org.uk/en/guidelines-research-services/coronavirus-covid-19-pregnancy-and-womens-health/ [accessed 2021-07-29]
2. Coronavirus disease (COVID-19): pregnancy and childbirth. World Health Organization. 2020. URL: https://www.who.int/news-room/q-a-detail/q-a-on-COVID-19-pregnancy-childbirth-and-breastfeeding [accessed 2021-07-29]
3. Poon L, Yang H, Kapur A, Melamed N, Dao B, Divakar H, et al. Global interim guidance on coronavirus disease 2019 (COVID-19) during pregnancy and puerperium from FIGO and allied partners: Information for healthcare professionals. Int J Gynaecol Obstet 2020 Jun;149(3):273-286. [doi: 10.1002/ijgo.13156] [Medline: 32248521]
4. Guidance for antenatal and postnatal services in the evolving coronavirus (COVID-19) pandemic. Royal College of Obstetrics and Gynaecology. 2020. URL: https://www.rcog.org.uk/globalassets/documents/guidelines/2020-07-10-guidance-for-antenatal-and-postnatal.pdf [accessed 2021-07-29]
5. Du L, Gu Y, Cui M, Li W, Wang J, Zhu L, et al. [Investigation on demands for antenatal care services among 2 002 pregnant women during the epidemic of COVID-19 in Shanghai]. Zhonghua Fu Chan Ke Za Zhi 2020 Mar 25;55(3):160-165. [doi: 10.3760/cma.j.cn112141-20200218-00112] [Medline: 32268713]

6. Aziz A, Zork N, Aubey JJ, Baptiste CD, D'Alton ME, Emeruwa UN, et al. Telehealth for high-risk pregnancies in the setting of the COVID-19 pandemic. Am J Perinatol 2020 Jun;37(8):800-808 [FREE Full text] [doi: 10.1055/s-0040-1712121] [Medline: 32396948]
7. Chapter 5: digitally-enabled care will go mainstream across the NHS. NHS Long Term Plan. 2019. URL: https://www.longtermplan.nhs.uk/online-version/chapter-5-digitally-enabled-care-will-go-mainstream-across-the-nhs/ [accessed 2020-06-26]
8. Pflugeisen BM, Mou J. Patient satisfaction with virtual obstetric care. Matern Child Health J 2017 Jul;21(7):1544-1551. [doi: 10.1007/s10995-017-2284-1] [Medline: 28176034]
9. 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]
10. Peahl AF, Smith RD, Moniz MH. Prenatal care redesign: creating flexible maternity care models through virtual care. Am J Obstet Gynecol 2020 Sep;223(3):389.e1-389.e10 [FREE Full text] [doi: 10.1016/j.ajog.2020.05.029] [Medline: 32425200]
11. Quinn L, Olajide O, Breslin E, Tan B, Ansar H. Virtual antenatal clinics. Br J Midwifery 2020 Sep 02;28(9):680-682. [doi: 10.12968/bjom.2020.28.9.680]
12. mHealth: New Horizons for Health Through Mobile Technologies. Geneva, Switzerland: World Health Organization; 2011:3.
13. Wallwiener S, Müller M, Doster A, Laserer W, Reck C, Pauluschke-Fröhlich J, et al. Pregnancy eHealth and mHealth: user proportions and characteristics of pregnant women using Web-based information sources-a cross-sectional study. Arch Gynecol Obstet 2016 Nov;294(5):937-944. [doi: 10.1007/s00404-016-4093-y] [Medline: 27084763]
14. 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]
15. Fryer K, Delgado A, Foti T, Reid CN, Marshall J. Implementation of obstetric telehealth during COVID-19 and beyond. Matern Child Health J 2020 Sep;24(9):1104-1110 [FREE Full text] [doi: 10.1007/s10995-020-02967-7] [Medline: 32564248]
16. Novick G. Women's experience of prenatal care: an integrative review. J Midwifery Womens Health 2009;54(3):226-237 [FREE Full text] [doi: 10.1016/j.jmwh.2009.02.003] [Medline: 19410215]
17. Sikorski J, Wilson J, Clement S, Das S, Smeeton N. A randomised controlled trial comparing two schedules of antenatal visits: the antenatal care project. BMJ 1996 Mar 02;312(7030):546-553 [FREE Full text] [doi: 10.1136/bmj.312.7030.546] [Medline: 8595286]
18. Holcomb D, Faucher M, Bouzid J, Quint-Bouzid M, Nelson D, Duryea E. Patient perspectives on audio-only virtual prenatal visits amidst the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic. Obstet Gynecol 2020 Aug;136(2):317-322. [doi: 10.1097/AOG.0000000000004026] [Medline: 32544144]
19. Futterman I, Rosenfeld E, Toaff M, Boucher T, Golden-Espinal S, Evans K, et al. Addressing disparities in prenatal care via telehealth during COVID-19: prenatal satisfaction survey in East Harlem. Am J Perinatol 2021 Jan;38(1):88-92 [FREE Full text] [doi: 10.1055/s-0040-1718695] [Medline: 33038898]
20. Butler Tobah YS, LeBlanc A, Branda ME, Inselman JW, Morris MA, Ridgeway JL, et al. Randomized comparison of a reduced-visit prenatal care model enhanced with remote monitoring. Am J Obstet Gynecol 2019 Dec;221(6):638.e1-638.e8. [doi: 10.1016/j.ajog.2019.06.034] [Medline: 31228414]
21. DeNicola N, Grossman D, Marko K, Sonalkar S, Butler Tobah YS, Ganju N, et al. Telehealth interventions to improve obstetric and gynecologic health outcomes: a systematic review. Obstet Gynecol 2020 Feb;135(2):371-382 [FREE Full text] [doi: 10.1097/AOG.0000000000003646] [Medline: 31977782]
22. Partridge CA, Holman JR. Effects of a reduced-visit prenatal care clinical practice guideline. J Am Board Fam Pract 2005;18(6):555-560 [FREE Full text] [doi: 10.3122/jabfm.18.6.555] [Medline: 16322418]
23. Villar J, Carroli G, Khan-Neelofur D, Piaggio G, Gülmezoglu M. Patterns of routine antenatal care for low-risk pregnancy. Cochrane Database Syst Rev 2001(4):CD000934. [doi: 10.1002/14651858.CD000934] [Medline: 11687086]
24. Dashraath P, Wong JLJ, Lim MXK, Lim LM, Li S, Biswas A, et al. Coronavirus disease 2019 (COVID-19) pandemic and pregnancy. Am J Obstet Gynecol 2020 Jun;222(6):521-531 [FREE Full text] [doi: 10.1016/j.ajog.2020.03.021] [Medline: 32217113]
25. Quinn LM, Davies MJ, Hadjiconstantinou M. Virtual consultations and the role of technology during the COVID-19 pandemic for people with type 2 diabetes: the UK perspective. J Med Internet Res 2020 Aug 28;22(8):e21609 [FREE Full text] [doi: 10.2196/21609] [Medline: 32716898]

Abbreviations

HCP: health care professional
NHS: National Health Service

Edited by G Eysenbach; submitted 05.11.20; peer-reviewed by K Fitzner, S Nabhani-Gebara; comments to author 30.11.20; revised version received 31.01.21; accepted 14.04.21; published 31.08.21.

Please cite as:
Quinn LM, Olajide O, Green M, Sayed H, Ansar H
Patient and Professional Experiences With Virtual Antenatal Clinics During the COVID-19 Pandemic in a UK Tertiary Obstetric Hospital: Questionnaire Study
J Med Internet Res 2021;23(8):e25549
URL: https://www.jmir.org/2021/8/e25549
doi:10.2196/25549
PMID:34254940

©Lauren Marie Quinn, Oluwafumbi Olajide, Marsha Green, Hazem Sayed, Humera Ansar. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 31.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Comparison of the Differences Between Web-Based and Traditional Questionnaire Surveys in Pediatrics: Comparative Survey Study

Heping Fang[1], MSc; Ruoling Xian[1], MSc; Zhuoying Ma[1], MSc; Mingyue Lu[1], MSc; Yan Hu[1], PhD

[1] Department of Child Health Care, Children's Hospital of Chongqing Medical University, Chongqing, China

Corresponding Author:
Yan Hu, PhD
Department of Child Health Care
Children's Hospital of Chongqing Medical University
No.136, 2nd Zhongshan Road
Yuzhong District
Chongqing,
China
Phone: 86 023 63622764
Email: hy420@126.com

Abstract

Background: A web-based survey is a novel method for data capture. Some studies have applied web-based surveys in pediatrics, but few of them have reported data on the differences between web-based and traditional questionnaire surveys.

Objective: The objective of our study was to evaluate the internal consistency of a web-based survey and compare it with a traditional questionnaire survey in pediatrics.

Methods: A convenience sample of caregivers was invited to participate in the survey on feeding patterns and their children's eating behaviors if their children were aged 2 to 7 years. A web-based survey and a traditional questionnaire survey were carried out between October 2018 and July 2019. A total of 1085 caregivers were involved in this study, and they were divided into the following three groups based on methods and sources: (1) web-based survey from a web source, (2) web-based survey from a hospital source, and (3) traditional questionnaire survey from a hospital source. The data were then compared and analyzed.

Results: A total of 735 caregivers participated in the web-based survey and 350 caregivers participated in the traditional questionnaire survey, and 816 cases were then included in the analyses after data processing. The effective rate of the web-based survey was 70.1% (515/735), and the completeness rate of the traditional questionnaire survey was 86.0% (301/350). There were no significant differences between web-based surveys from different sources. However, demographic characteristics were significantly different between the web-based and traditional questionnaire surveys, mainly in terms of age and caregivers (χ^2_4=16.509, P=.002 and χ^2_4=111.464, P<.001, respectively). Caregivers of children aged 2 to 3 years and grandparents were more likely to respond to the web-based survey. Age-specific stratified analysis showed that the score of "monitoring" and the reporting rate of "poor appetite" in children aged 2 to 3 years were significantly higher in the web-based survey compared to the traditional questionnaire survey after adjusting for demographic characteristics.

Conclusions: A web-based survey could be a feasible tool in pediatric studies. However, differences in demographic characteristics and their possible impacts on the results should be considered in the analyses.

(J Med Internet Res 2021;23(8):e30861) doi:10.2196/30861

KEYWORDS

pediatrics; survey; questionnaire; web survey; comparative study

Introduction

A questionnaire survey is an important method in social science studies. The method was invented in the 1930s and has been widely adopted in psychological studies. With the advent of technology, a web-based survey has been appreciated as a novel method and has been applied in clinical studies [1]. To date, the number of "netizens" (citizens of the internet) in China has reached 989 million, with a 70.4% rate of internet penetration [2], which has provided an effective platform for web-based surveys. In addition, the development of mobile internet technologies has resulted in flexible web-based surveys due to mobile phone apps. The technology has a simple design and production, and has a growing frequency of usage.

Due to convenience, there has been more adoption of web-based surveys in large-scale studies. Although a larger sample size is associated with more statistical confidence of the results, this depends on the requirements and specifications of the survey design. Thus, robust study designs determine the web-based survey results, especially in the medical and health fields [3]. Although previous studies have shown that the accuracy of a web-based survey may be higher, it is primarily used in basic information collection, satisfaction evaluation, and marketing [4]. In addition, findings from fields related to subjective cognition have been shown to be highly vulnerable to survey methods [5]. In pediatrics, the participants of questionnaire surveys are mainly caregivers, who are often subjective; thus, the adopted survey methods might yield varying data. While some studies have applied web-based surveys in pediatrics, data on the differences between web-based and traditional questionnaire surveys remain scant. Here, we evaluated the internal consistency of a web-based survey and then compared it with a traditional questionnaire survey in pediatrics.

Methods

Study Design

This study was conducted between October 2018 and July 2019. We used two survey methods to investigate eating behaviors in children aged 2 to 7 years, as well as the feeding patterns of their caregivers. We conducted a web-based survey in October 2018, while a traditional questionnaire survey was conducted in July 2019. Based on the methods and sources, the participants were divided into the following three groups: (1) web-based survey from a web source (group A), (2) web-based survey from a hospital source (group B), and (3) traditional questionnaire survey from a hospital source (group C).

The study was reviewed and approved by the Ethics Committee of the Children's Hospital of Chongqing Medical University (2019-202).

Traditional Questionnaire Survey

The traditional questionnaire survey was mainly conducted by the investigators through face-to-face questioning in the child health care clinic, and informed consent was obtained orally before the investigation. The completed questionnaires were then collected and kept by the investigators. The questionnaire (43 items) included basic information about the children and the caregivers, and required the caregivers to evaluate the children's performance in the previous month using the Likert scoring method [6].

The simplified Child Feeding Questionnaire (CFQ) revised by Lixia et al [7] was used to assess the feeding patterns of the caregivers. The content included "monitoring," "pressure," and "restriction," each of which involved four items, making a total of 12 items. The score for each dimension was expressed as an average score for the four items in that dimension.

On the other hand, the simplified Chinese version of Identification and Management of Feeding Difficulties (IMFeD) [8] was used to investigate the children's eating behaviors. The content included five aspects of eating behavior problems, including "poor appetite" (four items), "food preference" (three items), "poor eating habit" (four items), "parental misperception" (two items), and "fear of feeding" (two items). An eating behavior problem existed when the score was lower than 40% of the total in each aspect.

Web-Based Survey

The web-based survey was set as an open survey. It was built by "wjx" (an online questionnaire platform in mainland China) and was promoted remotely through WeChat on the internet (web source), as well as a paper poster (Multimedia Appendix 1) with a QR code in the child health care clinic (hospital source). Initial contact with potential participants was made on the internet, and the questionnaire could be accessed directly through WeChat. The questionnaire contained five pages, and a brief introduction of the questionnaire was shown on the first page. The main content and scoring methods for the web-based survey were similar to the traditional survey, and basic information was collected by adaptive questioning. There were, however, three additional items in the web-based survey that were used to distinguish between (1) the web and hospital source, (2) healthy and sick children, and (3) high and poor quality of the response (self-assessment), resulting in a total of 46 items. The survey time was set at about 10 minutes. To ensure the honesty of the participants, survey participation was anonymous and voluntary. To ensure participation, feedback and suggestions would be shared at the end of the survey. To reduce repeated responses, each WeChat account of the survey could only send a single response. In addition, to reduce the possibility of misunderstanding, 10 caregivers were invited to adjust the expressions through pretests prior to the survey. Finally, the participants were required to complete all items before submitting the survey, and the answers could not be reviewed or changed.

Data Processing

Only investigators had access to the data. Data, such as IP address, phone number, and children's and caregivers' basic information, were recorded in Excel 2016 (Microsoft Corp). Then, the data were checked for duplication, and cases with a high degree of overlap (if any of the three aspects mentioned above were the same) were considered as repeated responses. Because psychological theory appreciates the first impression as more accurate [9], late responses, which were repetitive, were excluded. Finally, cases of sick children (eg, chronic liver or kidney diseases, developmental diseases, and food allergy) and poor quality of responses (ie, the participants themselves thought the answers were not rigorous) in the web-based survey were also excluded from the analyses.

Statistical Analyses

The sample in this study was a convenience sample, and the sample size was calculated using a single proportion sample size estimating algorithm [10]. An 82.8% rate of incidence of eating behavior problems was taken from our previous study

[11], with a margin of error of 5%, confidence level of 95%, and nonresponse rate of 10%. Calculation provided 217 as the minimum sample size for each group.

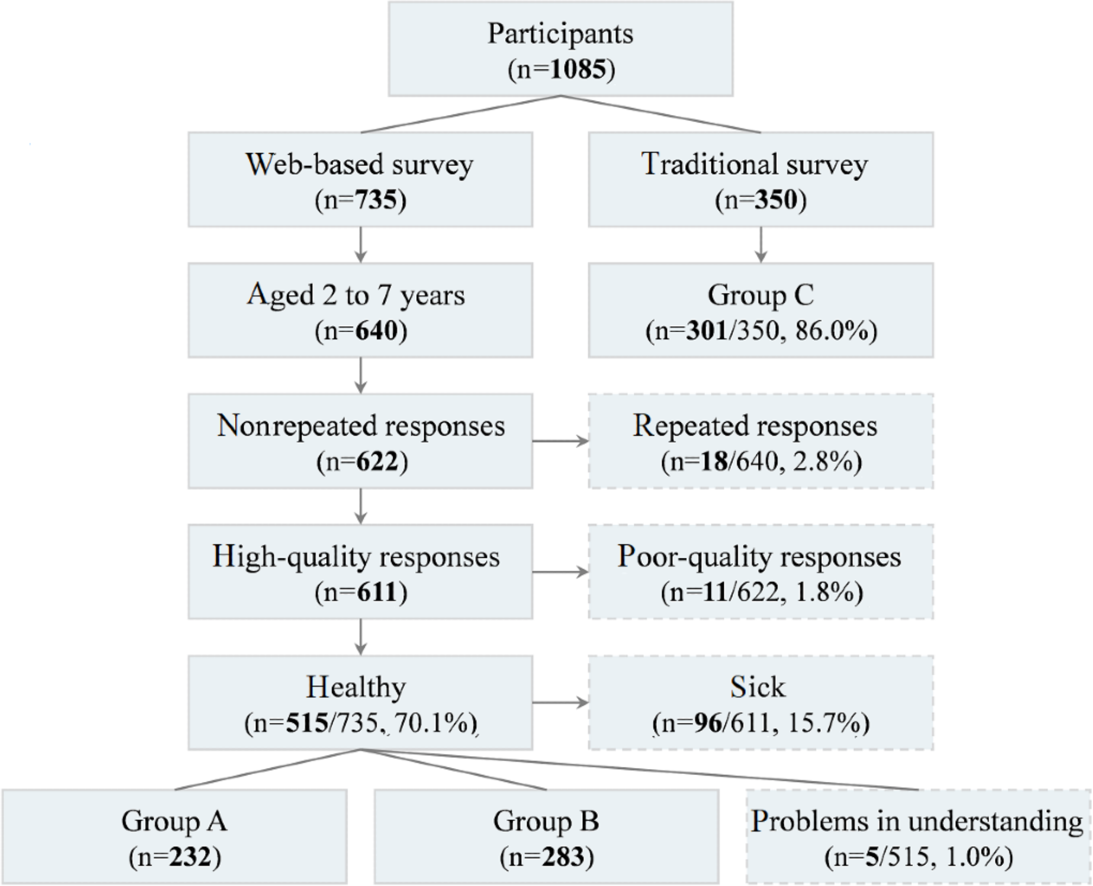
Only completed questionnaires were analyzed. All statistical analyses were performed using SPSS 26.0 (IBM Corp). Descriptive statistical tools were used to analyze the demographic characteristics of the participants. To test whether the children from the hospital sources (groups B and C) had nutrition-related problems, comparison between BMI-for-age Z scores and the standard mean was carried out using a one-sample *t* test. We used analysis of variance (ANOVA) and the least significant difference (LSD) test to compare the feeding patterns. On the other hand, comparison of eating behavior problems was carried out using the chi-square test. Bonferroni correction was utilized for multiple comparisons. A *P* value <.05 was considered statistically significant.

Results

Data Processing and Response Rates

Data processing and response rates are shown in Figure 1. A total of 1085 caregivers (735 were in the web-based survey and 350 were in the traditional questionnaire survey) were involved in this study. It took 9.9 (IQR 4.9) minutes for the web-based survey and about 10 minutes for the traditional questionnaire survey. There were 232 cases in group A, 283 in group B, and 301 in group C after data processing. The effective rate of the web-based survey was 70.1% (515/735), and the completeness rate of the traditional questionnaire survey was 86.0% (301/350).

Figure 1. Data processing and the response rates of the surveys.


Participants
(n=**1085**)
Web-based survey
(n=**735**)
Traditional survey
(n=**350**)
Aged 2 to 7 years
(n=**640**)
Group C
(n=**301**/350, 86.0%)
Nonrepeated responses
(n=**622**)
Repeated responses
(n=**18**/640, 2.8%)
High-quality responses
(n=**611**)
Poor-quality responses
(n=**11**/622, 1.8%)
Healthy
(n=**515**/735, 70.1%)
Sick
(n=**96**/611, 15.7%)
Group A
(n=**232**)
Group B
(n=**283**)
Problems in understanding
(n=**5**/515, 1.0%)

Characteristics of the Participants

Our data showed that the mean BMI-for-age Z scores in groups B and C were −0.06 (SD 1.08) and −0.09 (SD 1.31), respectively, which had no significant differences with the standard mean (t_{237}=−0.890, *P*=.37 and t_{271}=−1.133, *P*=.26, respectively). Thus, the children from hospital sources had no nutrition-related problems.

The chi-square test showed significant differences in age and caregivers among the three groups (χ^2_4=16.509, *P*=.002 and χ^2_4=111.464, *P*<.001, respectively). In addition, multiple comparisons showed that there were no significant differences between groups A and B. On the contrary, the results showed that there were significant differences between groups A and C in age and caregivers (χ^2_2=14.218, *P*=.001 and χ^2_2=103.387, *P*<.001, respectively), as well as between groups B and C in age and caregivers (χ^2_2=9.359, *P*=.009 and χ^2_2=96.940, *P*<.001, respectively). The findings showed that caregivers of children aged 2 to 3 years and grandparents were more likely to respond to the web-based survey (Table 1).

Table 1. Comparison of demographic characteristics among the three groups.

Characteristic	Group A[a] (N=232), n (%)	Group B[b] (N=283), n (%)	Group C[c] (N=301), n (%)	χ^2 (*df*)[d]	*P* value
Age				16.509 (4)	.002
2-3 years	106 (45.7%)	117 (41.4%)	90 (29.9%)[e,f]		
3-5 years	87 (37.5%)	119 (42.0%)	141 (46.8%)		
5-7 years	39 (16.8%)	47 (16.6%)	70 (23.3%)		
Gender				1.297 (2)	.52
Male	122 (52.6%)	163 (57.6%)	167 (55.5%)		
Female	110 (47.4%)	120 (42.4%)	134 (44.5%)		
Caregivers				111.464 (4)	<.001
Parents	146 (62.9%)	185 (65.4%)	290 (96.3%)[e,f]		
Grandparents	79 (34.1%)	90 (31.8%)	6 (2.0%)		
Others	7 (3.0%)	8 (2.8%)	5 (1.7%)		

[a] Group A: web-based survey from a web source.

[b] Group B: web-based survey from a hospital source.

[c] Group C: traditional questionnaire survey from a hospital source.

[d] Chi-square analysis with Bonferroni correction.

[e] *P*<.01 vs group A.

[f] *P*<.01 vs group B.

Feeding Patterns and Eating Behavior Problems

Age-specific stratified analysis was performed following selection of children whose caregivers were parents to avoid the influence of demographic characteristics on the data. The results showed that in children aged 2 to 3 years, the scores on “monitoring” and the rates of reporting “poor appetite” were significantly different among the three groups (F_2=12.549, *P*<.001 and χ^2_2=6.579, *P*=.04, respectively). Multiple comparisons then showed that the scores on “monitoring” in groups A and B were significantly higher compared to that in group C (*P*<.001 and *P*<.001, respectively), while the rate of reporting “poor appetite” in group B was significantly higher than that in group C (χ^2_1=6.138, *P*=.01). In children aged 3 to 5 years, the scores on “restriction” were significantly different among the three groups (F_2=3.221, *P*=.04), but multiple comparisons showed no significant differences between the groups. In children aged 5 to 7 years, the rates of reporting “poor appetite” were significantly different among the three groups (χ^2_2=6.472, *P*=.04), but multiple comparisons showed no significant differences between the groups.

Discussion

Principal Findings

Compared with a traditional questionnaire survey, a web-based survey is widely accepted due to its convenience, low cost, efficiency, wide-range coverage, and semiclosed and anonymous nature. However, it presents other shortcomings, such as repeated responses as well as information and selection bias, which might significantly affect the results of descriptive studies [12-14]. It is, therefore, necessary to conduct appropriate pretests [15] and choose reliable scientific data processing methods before executing a web-based survey. In our study, a variety of methods, such as pretests and outcome feedback, were adopted to improve the robustness of the web-based survey. Our results showed that only 2.8% (18/640) of the cases were classified as repeated responses, while 1.8% (11/622) were of poor quality. Moreover, the effective rate of the web-based survey was 70.1% (515/735), which was lower than the completeness rate of the traditional questionnaire survey (86.0%, 301/350). This is because the web-based survey was set as an open and voluntary survey, resulting in 95 participants (12.9%, 95/735) who were not study participants. These findings suggest that a web-based survey could be a feasible tool in pediatrics. Most importantly, the different data sources did not impact the output of the web-based survey, indicating that the web-based survey had good internal consistency whether through web promotion or hospital publicity. However, despite revising the questions based on the pretests, 1.0% (5/515) of the cases reported difficulty with understanding the content. Thus, lack of investigator explanation might lead to confusion of the results due to misunderstanding among participants.

We next assessed whether the different sources of data could impact the results. Benedik et al [16] showed that both web-based and traditional questionnaire surveys demonstrated no differences in nutrient intake among pregnant women. Although the study ensured the consistency of participants, highly educated participants may make the conclusion unrepresentative. In addition, since nutrient intake represents objective data, it may not be easily affected by the survey methods. On the contrary, Milton et al [17] showed that the web-based survey had a higher reporting rate on sensitive topics in young people, suggesting that the results between the two methods of surveys might differ even in cases where the participants are the same. The effects of web-based and traditional survey methods among different populations have also been evaluated. Graefe et al [18] reported different demographic characteristics between web-based and mail-based surveys, but the participants of the web-based survey were younger and had better education. In our study, caregivers of children aged 2 to 3 years and grandparents were more likely to participate in the web-based survey, which is different from the finding of other studies [18,19] reporting that older people are less likely to participate in a web-based survey. This may be explained by the fact that today's grandparents are not exactly the same as older people, and the widespread popularity of the internet in China [2] has undoubtedly changed how people use mobile phones. More importantly, different demographic characteristics might lead to different results. For instance, Cantuaria et al [19] demonstrated that participants in a mail-based survey were more likely to report health problems. In our study, we showed that not only the eating behaviors of children but also the feeding patterns of caregivers were inconsistent, even after adjusting for demographic characteristics, a phenomenon dominant in the younger age group. We associated these findings with the fact that our survey focused on children's behavioral development, and the answers were more subjective and more dependent on the feelings of the parents. Thus, a web-based survey might be more prone to information bias in pediatrics when it comes to subjective topics. Besides, the interpretations of the investigators might play a certain role in the traditional questionnaire survey and the psychological effects of the participants may influence them to give a "good" answer [20]. In addition, although the web-based survey was meant to be a self-evaluation model to improve participation, the potential selection effect of its voluntary nature on participants could not be completely ruled out, which might lead to an increased proportion of problematic participants. Taken together, our data suggest that demographic characteristics and results would be different among different survey methods when used in pediatrics, and topics with strong subjectivity need more comprehensive consideration before adopting the study tools.

Limitations

Our study was limited by the convenience sample, as well as the fact that the interval between the two surveys was slightly long and the participants were not the same. Although stratified analysis was conducted according to demographic characteristics, changes in the feeding and eating behaviors of the participants and their effects on the results cannot be completely ruled out.

Conclusions

Our data demonstrated that a web-based survey could be a feasible tool in pediatrics. However, differences in demographic characteristics and their possible impacts on the results should be considered when interpreting data from a web-based survey.

Acknowledgments

We thank all the participants in the study and the members of the survey team, as well as the Children's Hospital of Chongqing Medical University. We also thank Ms Zhiyi Chen for designing the paper poster.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The paper poster used in the survey.
[PNG File , 4969 KB - jmir_v23i8e30861_app1.png]

References

1. Palmblad M, Tiplady B. Electronic diaries and questionnaires: Designing user interfaces that are easy for all patients to use. Qual Life Res 2004 Sep;13(7):1199-1207. [doi: 10.1023/B:QURE.0000037501.92374.e1] [Medline: 15473498]
2. The 47th China statistical report on internet development. China INIC. URL: http://www.cac.gov.cn/2021-02/03/c_1613923423079314.htm [accessed 2021-02-03]
3. Knitza J, Vossen D, Geffken I, Krusche M, Meyer M, Sewerin P, Arbeitskreis Junge Rheumatologen. [Use of medical apps and online platforms among German rheumatologists : Results of the 2016 and 2018 DGRh conference surveys and research conducted by rheumadocs]. Z Rheumatol 2019 Nov;78(9):839-846. [doi: 10.1007/s00393-018-0578-3] [Medline: 30542914]
4. Minto C, Vriz GB, Martinato M, Gregori D. Electronic Questionnaires Design and Implementation. Open Nurs J 2017;11:157-202 [FREE Full text] [doi: 10.2174/1874434601711010157] [Medline: 29238422]
5. Braekman E, Charafeddine R, Demarest S, Drieskens S, Berete F, Gisle L, et al. Comparing web-based versus face-to-face and paper-and-pencil questionnaire data collected through two Belgian health surveys. Int J Public Health 2020 Jan;65(1):5-16. [doi: 10.1007/s00038-019-01327-9] [Medline: 31993674]
6. Ventegodt S, Merrick J, Andersen NJ. Measurement of quality of life II. From the philosophy of life to science. ScientificWorldJournal 2003 Oct 13;3:962-971 [FREE Full text] [doi: 10.1100/tsw.2003.76] [Medline: 14570987]
7. Lixia Z, Daoping S, Chulin C, Fangfang L, Daqiao Z. Reliability and validity of a Chinese version of Child Feeding Questionnaire among parents of preschoolers in Chinese. Chinese Journal of Child Health Care 2016 Oct;24(10):1019-1023. [doi: 10.11852/zgetbjzz2016-24-10-04]

8. Jing L, Yanhong H, Ying D, Chunlan C, Lili L, Xuejiao Z, et al. Intervention effects evaluation about the eating behavior problems of children by Identification and Management of Feeding Difficulties tools, China in Chinese. Chinese Journal of Child Health Care 2014 Mar;22(03):333-336 [FREE Full text]
9. Biesanz JC, Human LJ, Paquin A, Chan M, Parisotto KL, Sarracino J, et al. Do We Know When Our Impressions of Others Are Valid? Evidence for Realistic Accuracy Awareness in First Impressions of Personality. Social Psychological and Personality Science 2011 Jan 19;2(5):452-459. [doi: 10.1177/1948550610397211]
10. Pourhoseingholi MA, Vahedi M, Rahimzadeh M. Sample size calculation in medical studies. Gastroenterol Hepatol Bed Bench 2013;6(1):14-17 [FREE Full text] [Medline: 24834239]
11. Mingyue L. The effects of feeding styles on eating behavior and BMI of children aged 2 to 7. China National Knowledge Infrastructure (CNKI). URL: https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD201901&filename=1018864932.nh [accessed 2018-05-01]
12. Bohannon J. SCIENTIFIC INTEGRITY. Survey fraud test sparks battle. Science 2016 Mar 04;351(6277):1014. [doi: 10.1126/science.351.6277.1014] [Medline: 26941296]
13. Alessi EJ, Martin JI. Conducting an Internet-based Survey: Benefits, Pitfalls, and Lessons Learned. Social Work Research 2010 Jun 01;34(2):122-128. [doi: 10.1093/swr/34.2.122]
14. Hohwü L, Lyshol H, Gissler M, Jonsson SH, Petzold M, Obel C. Web-based versus traditional paper questionnaires: a mixed-mode survey with a Nordic perspective. J Med Internet Res 2013 Aug 26;15(8):e173 [FREE Full text] [doi: 10.2196/jmir.2595] [Medline: 23978658]
15. Ball HL. Conducting Online Surveys. J Hum Lact 2019 Aug;35(3):413-417. [doi: 10.1177/0890334419848734] [Medline: 31084575]
16. Benedik E, Koroušić Seljak B, Simčič M, Rogelj I, Bratanič B, Ding EL, et al. Comparison of paper- and web-based dietary records: a pilot study. Ann Nutr Metab 2014;64(2):156-166. [doi: 10.1159/000363336] [Medline: 25116257]
17. Milton AC, Ellis LA, Davenport TA, Burns JM, Hickie IB. Comparison of Self-Reported Telephone Interviewing and Web-Based Survey Responses: Findings From the Second Australian Young and Well National Survey. JMIR Ment Health 2017 Sep 26;4(3):e37 [FREE Full text] [doi: 10.2196/mental.8222] [Medline: 28951382]
18. Graefe A, Mowen A, Covelli E, Trauntvein N. Recreation Participation and Conservation Attitudes: Differences Between Mail and Online Respondents in a Mixed-Mode Survey. Human Dimensions of Wildlife 2011 May 31;16(3):183-199. [doi: 10.1080/10871209.2011.571750]
19. Cantuaria ML, Blanes-Vidal V. Self-reported data in environmental health studies: mail vs. web-based surveys. BMC Med Res Methodol 2019 Dec 12;19(1):238 [FREE Full text] [doi: 10.1186/s12874-019-0882-x] [Medline: 31830906]
20. Phillips DL, Clancy KJ. Some Effects of "Social Desirability" in Survey Studies. American Journal of Sociology 1972 Mar;77(5):921-940. [doi: 10.1086/225231]

Edited by R Kukafka; submitted 01.06.21; peer-reviewed by M Hawkins, R Watson; comments to author 28.06.21; revised version received 18.07.21; accepted 27.07.21; published 26.08.21.

Please cite as:
Fang H, Xian R, Ma Z, Lu M, Hu Y
Comparison of the Differences Between Web-Based and Traditional Questionnaire Surveys in Pediatrics: Comparative Survey Study
J Med Internet Res 2021;23(8):e30861
URL: https://www.jmir.org/2021/8/e30861
doi:10.2196/30861
PMID:34319240

©Heping Fang, Ruoling Xian, Zhuoying Ma, Mingyue Lu, Yan Hu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Health Information Technology Use Among Persons With Self-reported Atherosclerotic Cardiovascular Disease: Analysis of the 2011-2018 National Health Interview Survey

Uchenna Nwokeji[1], MS; Erin M Spaulding[2,3,4], RN, BSN, PhD; Rongzi Shan[5], MD; Ruth-Alma Turkson-Ocran[6], FNP-BC, RN, MPH, PhD; Diana Baptiste[2], CNE, RN, DNP; Binu Koirala[2], RN, MGS, PhD; Timothy B Plante[7], MHS, MD; Seth S Martin[3,4,8,9], MHS, MD; Yvonne Commodore-Mensah[2,4,9], RN, MHS, PhD

[1]Department of Pharmacology and Cellular Biology, University of Pittsburgh School of Medicine, Pittsburgh, PA, United States

[2]Johns Hopkins University School of Nursing, Baltimore, MD, United States

[3]The Welch Center for Prevention, Epidemiology and Clinical Research, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, United States

[4]Center for Mobile Technologies to Achieve Equity in Cardiovascular Health, American Heart Association Strategically Focused Research Network Health Technology and Innovation Center, Baltimore, MD, United States

[5]Department of Medicine, David Geffen School of Medicine, Los Angeles, CA, United States

[6]Division of General Internal Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, United States

[7]Department of Medicine, Larner College of Medicine at the University of Vermont, Burlington, VT, United States

[8]Ciccarone Center for the Prevention of Cardiovascular Disease, Division of Cardiology, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, United States

[9]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

Corresponding Author:
Yvonne Commodore-Mensah, RN, MHS, PhD
Department of Epidemiology
Johns Hopkins Bloomberg School of Public Health
525 N. Wolfe Street, Room #419
Baltimore, MD, 21205
United States
Phone: 1 410 614 1529
Email: ycommod1@jhu.edu

Abstract

Background: Atherosclerotic cardiovascular disease (ASCVD) is the leading cause of morbidity and mortality in the United States. Health information technologies (HITs) have recently emerged as a viable intervention to mitigate the burden of ASCVD. Approximately 60% of US adults report searching the internet for health information; however, previous research has not examined the prevalence of general technology or HIT use among adults with and without ASCVD. In addition, social determinants in HIT use among adults with ASCVD are not well understood.

Objective: The aim of this study was to evaluate the prevalence and social determinants of HIT use among US adults with versus without self-reported ASCVD.

Methods: We pooled cross-sectional data from the 2011-2018 National Health Interview Survey (NHIS) to examine the general technology and HIT use among adults aged ≥18 years with and without self-reported ASCVD (coronary heart disease, stroke, or both). General technology use was defined as mobile phone ownership, internet use, and computer use. HIT use was defined as looking up health information on the internet, filling a web-based prescription, scheduling a medical appointment on the internet, communicating with a health care provider by email, or using web-based group chats to learn about health topics. We evaluated sociodemographic differences in HIT use among respondents by using Poisson regression. Analyses were weighted according to NHIS standards.

Results: A total sample of 256,117 individuals were included, of which 2194 (0.9%) reported prior ASCVD. Among adults with prior ASCVD, the mean age was 70.6 (SD 11.5) years, and 47.4% (1048/2194) of the adults were females. General technology use differed between participants with and without prior ASCVD, with 36.0% (614/1826) and 76.2% (157,642/213,816) indicating internet usage and 24.6% (374/1575) and 60.7% (107,742/184,557) indicating using a computer every day, respectively. Similarly, adults with ASCVD were less likely to use HIT than those without ASCVD (515/2194, 25.1% vs 123,966/253,923, 51.0%; *P*<.001). Among adults with prior ASCVD, social determinants that were associated with HIT use included younger age, higher education, higher income, being employed, and being married.

Conclusions: HIT use was low among adults with a history of ASCVD, which may represent a barrier to delivering care via emerging HIT. Given the associations with social determinants such as income, education, and employment, targeted strategies and policies are needed to eliminate barriers to impact HIT usage.

(J Med Internet Res 2021;23(8):e23765) doi:10.2196/23765

KEYWORDS

health information technology; cardiovascular disease; digital health; eHealth; mobile phone

Introduction

Atherosclerotic cardiovascular disease (ASCVD) is the leading cause of morbidity and mortality in the United States, accounting for more lives lost per year than those lost to cancer and chronic lung disease combined [1]. In 2016, over 360,000 people died from coronary heart disease, which is the most prevalent form of heart disease. Additionally, 1 in every 19 deaths in the United States, on average, was due to stroke. From 2014 to 2015, the economic burden associated with ASCVD was estimated to result in US $351.2 billion in direct and indirect annual costs [1]. Adults with prior ASCVD events are at high risk for recurrence and require contemporary secondary prevention strategies, including behavioral and medical interventions [2]. ASCVD outcomes are also disparate among groups of persons with different social determinants, including persons with lower income, lower education, and racial minority status [3]. Novel technologies may play a role in addressing the barriers contributing to disparities in ASCVD outcomes.

Health information technologies (HITs) have recently emerged as a viable intervention to mitigate the burden of ASCVD [4,5]. HITs, which encompass patient portals, mobile phone interventions, electronic health records, and telemedicine services, are increasingly being used to improve communication between patients and clinicians and to facilitate chronic disease management [6,7]. Approximately 60% of US adults report searching the internet for health information [8]. Around late March/early April 2020, the implementation of telemedicine services to meet patient demand drastically increased in response to the COVID-19 pandemic, which has reached 92,262,621 cases in 223 countries as of January 16, 2021 [9-11]. Historically, however, there has been a "digital divide" in which underserved populations lack access to computers and the internet, thereby serving as a significant barrier to care management utilizing HIT [12]. In recent years, mobile phones have helped in bridging this divide. In terms of options for online access, approximately 25% of Hispanics and 23% of Non-Hispanic Blacks are "smartphone only" internet users in place of traditional home broadband services compared to only 12% of Whites [13]. Non-Hispanic Blacks and Hispanics are also more likely than Non-Hispanic Whites to seek health information via their smartphones [8].

Previous studies on the sociodemographic characteristics of general US adult population using HIT have found that these adults tend to be Whites, women, young, and have a higher income and education level [14-16]. In terms of HIT use among the CVD population, 73% of US adults with or at risk for CVD owned a smartphone and 48% had a health app [17]. Adults with or at risk for CVD were also more likely to share health information from a smartphone/wearable device with a clinician [17]. Few studies utilizing nationally representative databases have assessed sociodemographic differences in HIT use among adults with self-reported ASCVD. To better understand the differences in HIT use among adults with ASCVD, we analyzed the National Health Interview Survey (NHIS). Specifically, we sought to (1) evaluate general technology and HIT use by prior ASCVD status, (2) describe changes in HIT use over time, and (3) describe social determinants of HIT use among US adults with ASCVD. We hypothesized that among adults with prior ASCVD, (1) general technology and HIT use would be lower, (2) HIT use would increase over time, and (3) social determinants of health that are indicative of more vulnerable status would be associated with lower HIT use.

Methods

Study Population

Analyses were performed with cross-sectional data from the NHIS, a civilian noninstitutionalized population survey of US adults aged ≥18 years, which was administered by the National Center for Health Statistics (NCHS). From 2006 to 2015, NHIS employed a multistage stratified cluster probability design that oversampled Black, Hispanic, and Asian people; however, oversampling was stopped in 2016 to adjust for the changes in the distribution of the US population since 2006. NHIS also reaches 35,000 households with about 87,500 persons annually [18]. For the sample adult questionnaire, one randomly selected adult per family is interviewed in person by a trained NCHS staff member who records the participant's self-reported information on health care access and utilization, health status, behavior, and other sociodemographic data [18]. A full description of the NHIS methodologies can be found elsewhere [19]. The data for the years 2011-2018 were pooled using NCHS guidelines to improve the accuracy of the estimates [18].

Participants

Respondents included in the analysis were US adults aged ≥18 years who answered "yes" or "no" to "Have you ever been told by a doctor or other health professional that you had coronary heart disease?" We also included US adults who answered "yes" or "no" to "Have you ever been told by a doctor or other health professional that you had a stroke?" We defined ASCVD as self-reported coronary heart disease, stroke, or both. Of the

XSL•FO RenderX

257,653 participants in NHIS, we excluded participants who did not provide data on education (n=1115), employment (n=142), and health status (n=279). The analytical sample included 256,117 participants.

Outcome Measurements

General Technology Use

Mobile phone ownership, internet use, and computer use were compared between US adults with and without ASCVD to assess the prevalence of general technology use. Mobile phone ownership was derived from persons who responded "≥1" to the question: "How many working cell phones do you or people in your family have?" Internet use was derived from persons who answered "yes" to the question: "Do you use the internet?" Computer use was derived from persons who answered, "Never or almost never," "Some days," "Most days," or "Every day" to the question: "How often do you use a computer?"

HIT Use

HIT use, the primary outcome, was defined as responding "yes" from the responses "yes," "no," "refused," "not ascertained," or "don't know" to a question regarding computer use in the prior 12 months to do one of the following: looking up health information on the internet, filling a web-based prescription, scheduling a medical appointment on the internet, communicating with health care providers through email, or using web-based group chats to learn about health topics.

Covariates

Covariates included self-reported age, sex, race/ethnicity, education level, employment status, health insurance status, health status, marital status, and income. Education level was recorded as ≤high school, some college, or ≥bachelor's degree. Health insurance status responses were categorized as covered and not covered. Health status was defined as reporting feeling better, worse, or about the same compared to last year. Income was measured by poverty income ratio, which is a variable calculated by the NCHS using the midpoint family income divided by the poverty level in dollars, corresponding to the US Census Bureau of the same survey year.

Statistical Analysis

Sample weights recommended by NCHS for the analytic years were used to adjust for the complex sample design [19]. We examined demographic characteristics by survey-weighted percentages among adults with and without a history of ASCVD. Weighted percentages were also calculated to measure the prevalence of general technology use and HIT use. Due to the complex sampling strategy, resulting in varying weights of individual observations, the percentage calculated by dividing the raw number of adults by the total n of the category of interest does not necessarily equal to the tabulated weighted percentage. Chi-square tests were used to examine differences in the HIT use categories, with *P* values <.05 deemed statistically significant. We examined the estimated prevalence of HIT use by using generalized linear models with a Poisson distribution and logarithmic link with linearized variance estimation. We also examined the adjusted predicted values and marginal effects of the primary outcome. The model was adjusted for age, sex, race/ethnicity, education level, employment status, health insurance status, health status, marital status, and income. Statistical analyses were performed with Stata version 16.0 SE (StataCorp LLC).

Results

Demographics of the Study Population

Of the 256,117 participants, 2194 (0.9%) reported prior ASCVD. Table 1 displays the prevalence of the demographic characteristics among the subpopulations of adults with and without a history of ASCVD. The mean age of the participants was 70.6 (SD 11.5) years, 47.4% (1048/2194) were females, and 55.1% (1270/2194) received no more than high school education. Compared to respondents without a history of ASCVD, adults with a history of ASCVD were also more often male, older, non-Hispanic White, had lower income, unemployed, less educated, uninsured, and at a worse overall health status from the previous year.

Table 1. Demographics of the adults with and without a history of atherosclerotic cardiovascular disease in the National Health Interview Survey.

Sociodemographic characteristics	Adults with ASCVD[a] (n=2194), %[b] (95% CI)	Adults without ASCVD (n=253,923), % (95% CI)
Sex		
Male	52.6 (50.1-55.1)	45.8 (45.6-46.1)
Female	47.4 (44.9-49.9)	54.2 (53.9-54.5)
Age (years)		
70+	56.0 (53.6-58.4)	15.8 (15.5-16.1)
60-69	27.5 (25.3-29.8)	15.9 (15.6-16.1)
50-59	12.2 (10.7-13.9)	17.2 (16.9-17.4)
40-49	3.1 (2.4-4.1)	15.7 (15.5-15.9)
30-39	1.1 (0.6-1.8)	16.6 (16.4-16.8)
18-29	0.1 (0.05-0.3)	18.9 (18.5-19.4)
Race/ethnicity		
Non-Hispanic White	70.7 (68.3-72.9)	68.6 (67.9-69.2)
Hispanic	9.4 (7.9-11.1)	13.0 (12.5-13.5)
Non-Hispanic Black	15.9 (14.3-17.7)	12.5 (12.1-13.0)
Non-Hispanic Asian	2.9 (2.2-3.7)	4.9 (4.7-5.1)
Non-Hispanic multiple races and other races	1.1 (0.7-1.8)	1.0 (0.8-1.1)
Poverty income ratio		
Below poverty level	22.5 (20.5-24.5)	15.2 (14.9-15.6)
Between 100% and 200% of poverty level	30.9 (28.8-33.2)	19.0 (18.7-19.3)
>200% above poverty level	46.6 (44.1-49.2)	65.8 (65.2-66.3)
Employment status		
Not employed	91.5 (89.9-92.8)	41.5 (41.1-41.9)
Employed	8.5 (7.2-10.1)	58.5 (58.1-58.9)
Marital status		
Not married	62.9 (60.5-65.3)	56.1 (55.7-56.6)
Married	37.1 (34.7-39.5)	43.9 (43.4-44.3)
Health status		
Better	19.6 (17.7-21.6)	18.4 (18.2-18.6)
Worse	26.3 (24.3-28.5)	8.5 (8.4-8.7)
About the same	54.1 (51.6-56.6)	73.1 (72.9-73.3)
Education level		
≤High school	55.1 (52.5-57.6)	37.2 (36.7-37.7)
Some college	28.0 (25.9-30.3)	31.1 (30.8-31.5)
≥Bachelor’s degree	16.9 (15.2-18.9)	31.7 (31.1-32.2)
Insurance coverage		
Not covered	97.4 (96.4-98.1)	87.9 (87.7-88.2)
Covered	2.6 (1.9-3.6)	12.1 (11.8-12.4)

[a] ASCVD: atherosclerotic cardiovascular disease.

[b] Survey-weighted percentages.

Prevalence of Web Access and Electronic Device Ownership

Approximately 76.7% (1682/2194) of the US adults with ASCVD and 89.1% (226,357/253,923) of the US adults without ASCVD reported their household owning at least one mobile phone. Figure 1 compares mobile phone ownership among US adults with and without a history of ASCVD. Of 2194 US adults, 1826 (83.2%) adults with self-reported ASCVD reported on whether they used the internet. Of the total respondents, 36.0% (614/1826) of adults with a history of ASCVD indicated internet use. Approximately 84.2% (213,816/253,923) of the US adults without ASCVD reported on whether they used the internet. Of the total respondents, 76.2% (157,642/213,816) of adults without a history of ASCVD indicated internet use. Figure 2 compares internet use among US adults with and without a history of ASCVD. Approximately 71.8% (1575/2194) of the US adults with self-reported ASCVD reported on how often they use a computer. Of the total respondents, 24.6% (374/1575) of adults with a history of ASCVD indicated that they use a computer every day. Approximately 72.7% (184,557/253,923) of the US adults without ASCVD reported on how often they use a computer. Of the total respondents, 24.6% (374/1575) of adults with a history of ASCVD indicated that they use a computer every day. Approximately 72.6% (184,557/253,923) of the US adults without ASCVD reported on how often they use a computer. Of the total respondents, 60.7% (107,742/184,557) of adults without a history of ASCVD indicated that they use a computer every day. Figure 3 compares the frequency of computer use among US adults with and without a history of ASCVD.

Figure 1. Mobile phone ownership among US adults with and without a history of atherosclerotic cardiovascular disease. ASCVD: atherosclerotic cardiovascular disease.

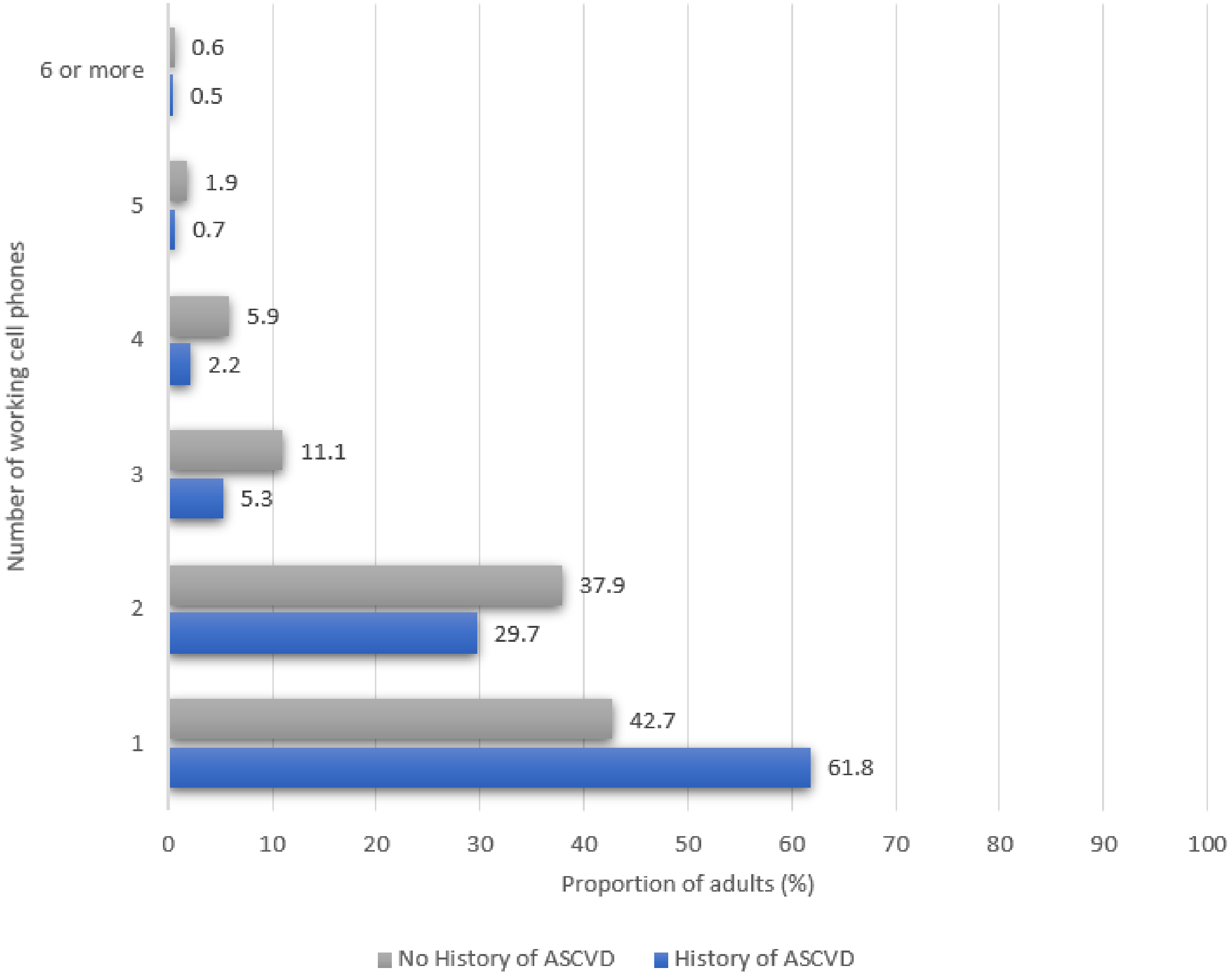


Figure 2. Internet use among US adults with and without a history of atherosclerotic cardiovascular disease. ASCVD: atherosclerotic cardiovascular disease.

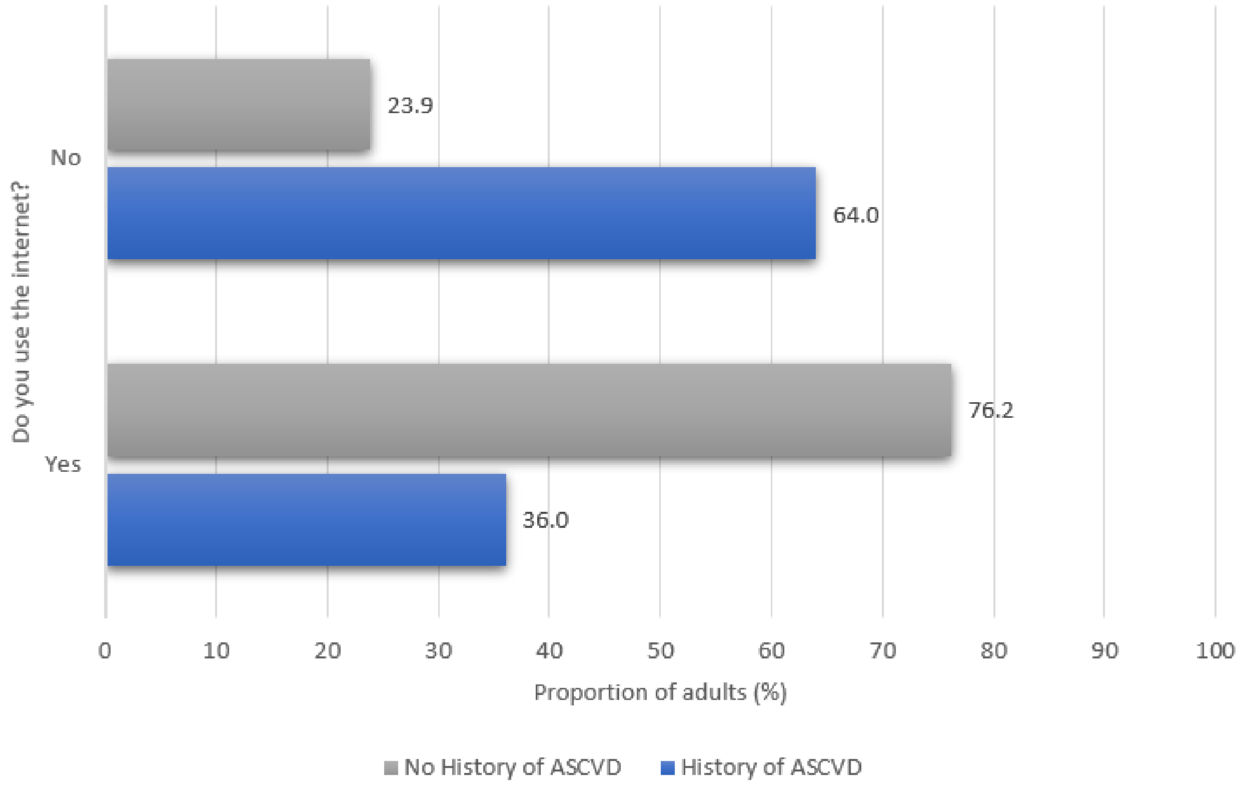
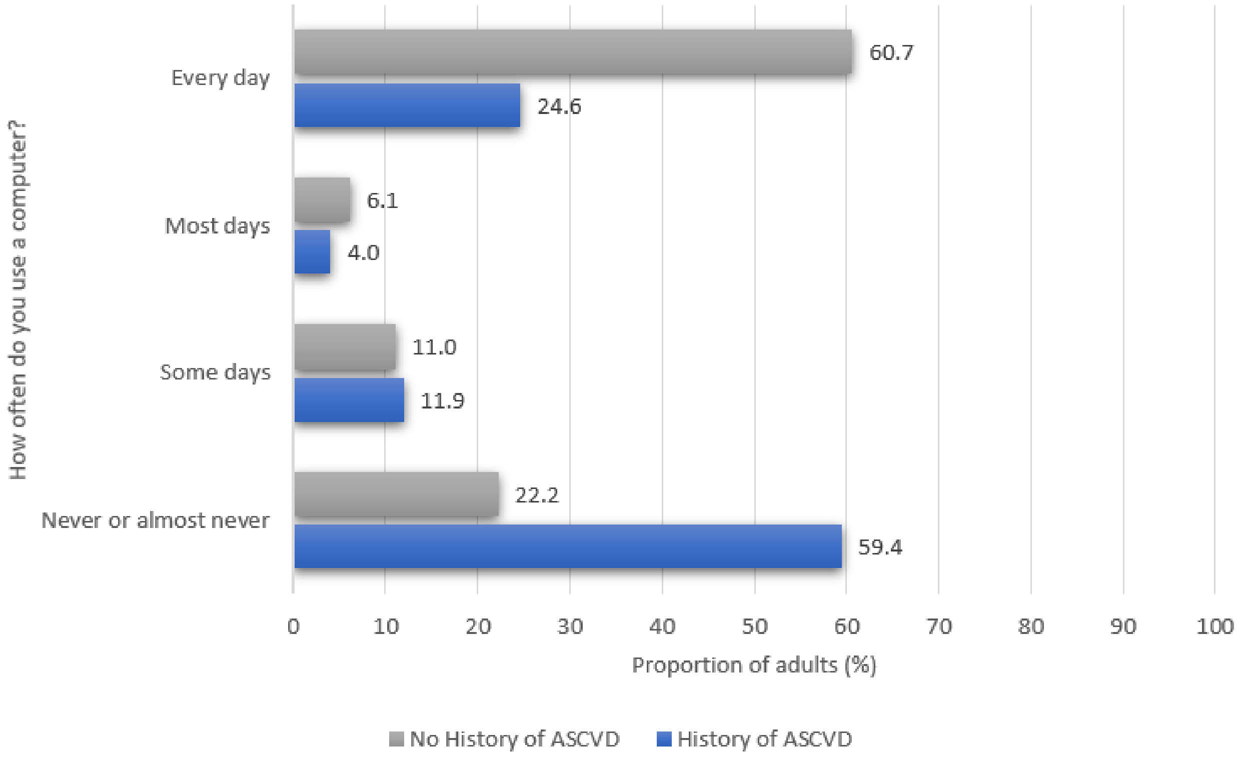


Figure 3. Frequency of computer use among US adults with and without a history of atherosclerotic cardiovascular disease. ASCVD: atherosclerotic cardiovascular disease.



HIT Use Differences Between Those With and Those Without ASCVD

Of the respondents with a history of ASCVD, 25.1% (515/2194) utilized some form of HIT (Table 2). Adults without a history of ASCVD utilized at least one form of HIT at a significantly higher proportion, that is, 51.0% (123,966/253,923). There were significant differences observed between adults with and without a history of ASCVD in 4 of the 5 individual indicators of HIT use (looking up health information on the internet, filling a web-based prescription, scheduling a medical appointment on the internet, communicating with a health care provider through email), with adults without a history of ASCVD utilizing these services more than adults with a history of ASCVD (Table 2).

Table 2. Weighted percentages and standard errors of health information technology use by adults with and without history of atherosclerotic cardiovascular disease.

Health information technology use indicators	All[a], % (SE)	Adults with ASCVD[b] (n=2194), % (SE)	Adults without ASCVD (n=253,923), % (SE)	*P* value
Overall health information technology use[c]	50.76 (0.2)	25.07 (1.1)	50.97 (0.2)	<.001
Looked up health information on the internet	48.72 (0.2)	23.13 (1.1)	48.94 (0.2)	<.001
Filled a web-based prescription	8.61 (0.1)	6.55 (0.6)	8.63 (0.1)	.004
Scheduled a medical appointment on the internet	9.53 (0.2)	4.70 (0.6)	9.57 (0.2)	<.001
Communicated with health care provider through email	10.38 (0.2)	5.39 (0.6)	10.42 (0.2)	<.001
Used online group chat to learn about health topics	3.50 (0.1)	3.35 (0.5)	3.50 (0.1)	.78

[a] Weighted percentage of all adults who used a form of health information technology.

[b] ASCVD: atherosclerotic cardiovascular disease.

[c] Use of at least one of the 5 indicators of health information technology.

Figure 4 compares the prevalence of overall HIT use among US adults with and without a history of ASCVD from 2011 to 2018. HIT use among respondents with a history ASCVD demonstrated an upward trend over this period, peaking in 2017 at 39.0% (81/229). US adults without a history of ASCVD exhibited higher overall HIT use prevalence compared to adults with a history of ASCVD each year over the time period, peaking in 2018 at 58.2% (14,571/25,163). Table 3 displays the unadjusted prevalence ratios for HIT use during 2011 to 2018. An increased likelihood of HIT use over time was demonstrated among US adults with and without a history of ASCVD, with a general increase over the years and the most recent years of 2017 (with ASCVD: 1.96, without ASCVD: 1.20*)* and 2018 (with ASCVD: 1.45, without ASCVD: 1.23) exhibiting the highest likelihood of use.

Figure 4. Trends in overall health information technology use among US adults with and without a history of atherosclerotic cardiovascular disease from 2011 to 2018. ASCVD: atherosclerotic cardiovascular disease; HIT: health information technology.

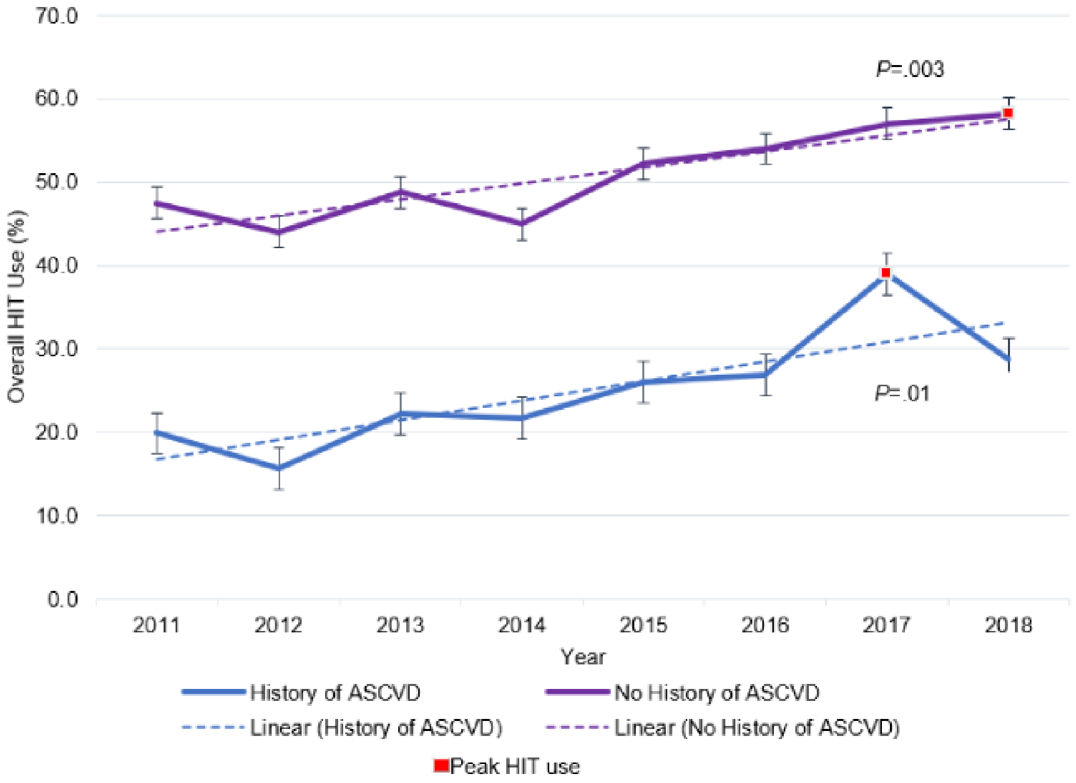


Table 3. Prevalence ratios for health information technology use over time among adults with and without a history of atherosclerotic cardiovascular disease.

Survey year	Adults with ASCVD[a] (n=2194), unadjusted PR[b] (95% CI)	*P* value	Adults without ASCVD (n=253,923), unadjusted PR (95% CI)	*P* value
2011	1.00 (Ref)[c]	N/A[d]	1.00 (Ref)	N/A
2012	0.78 (0.52-1.18)	.24	0.93 (0.91-0.95)	<.001
2013	1.11 (0.80-1.56)	.51	1.03 (1.01-1.05)	.004
2014	1.09 (0.77-1.53)	.63	0.95 (0.93-0.97)	<.001
2015	1.31 (0.94-1.82)	.11	1.10 (1.08-1.12)	<.001
2016	1.35 (0.96-1.89)	.08	1.14 (1.11-1.17)	<.001
2017	1.96 (1.45-2.65)	<.001	1.20 (1.17-1.23)	<.001
2018	1.45 (1.04-2.01)	.03	1.23 (1.20-1.25)	<.001

[a] ASCVD: atherosclerotic cardiovascular disease.

[b] PR: prevalence ratio.

[c] Ref: reference.

[d] N/A: not applicable.

Persons with a history of ASCVD who used HIT were more likely to be females (25.9%) and younger in age, as the 18-29 age group had the highest prevalence at 74.7%. Likewise, persons with a history of ASCVD aged ≥70 years had the lowest prevalence of overall HIT use at 19.6%. Among the racial/ethnic groups, non-Hispanic White adults had the highest prevalence of overall HIT use (26.4%) and non-Hispanic multiple race and other race adults had the lowest at 17.1%. Persons with a history of ASCVD who exhibited higher prevalence in overall HIT use were more likely to have a higher income being at least >200% above poverty level (27.4%), be employed (33.1%), be married (29.8%), and have a bachelor’s degree or greater (44.2%). Significant differences in the prevalence of overall HIT use among adults with a history of ASCVD were found in most of the social determinants evaluated and are detailed in Table 4.

Table 4. Sociodemographic characteristics of the adults with a history of atherosclerotic cardiovascular disease who used health information technology (n=2194).[a]

Sociodemographic characteristics	Used HIT[b], % (95% CI)	Unadjusted prevalence ratio (95% CI)	*P* value	Adjusted prevalence ratio[c] (95% CI)	*P* value
Gender					
Male	24.3 (21.5-27.0)	1.0 (Ref)[d]	N/A[e]	1.0 (Ref)	N/A
Female	25.9 (22.5-29.2)	0.9 (0.7-1.0)	.08	1.1 (0.9-1.3)	.40
Age (years)					
70+	19.6 (16.9-22.4)	1.0 (Ref)	N/A	1.0 (Ref)	N/A
60-69	29.2 (25.2-33.3)	*1.6 (1.3-1.9)*	<.001	*1.5 (1.3-1.8)*	<.001
50-59	33.3 (26.7-40.0)	*1.7 (1.3-2.2)*	<.001	*1.7 (1.3-2.2)*	<.001
40-49	42.2 (29.3-55.0)	*2.2 (1.6-3.1)*	<.001	*2.2 (1.6-3.1)*	<.001
30-39	46.1 (27.4-64.9)	*2.9 (1.9-4.6)*	<.001	*2.4 (1.6-3.7)*	<.001
18-29	74.7 (15.8-133.5)	*2.8 (1.1-7.1)*	.03	*3.9 (1.7-8.6)*	.001
Race/ethnicity					
Non-Hispanic White	26.4 (23.7-29.0)	1.0 (Ref)	N/A	1.0 (Ref)	N/A
Hispanic	23.4 (16.5-30.5)	*0.7 (0.5-1.0)*	.03	0.9 (0.7-1.2)	.43
Non-Hispanic Black	20.5 (15.9-25.1)	*0.7 (0.6-0.9)*	.01	*0.8 (0.6-1.0)*	.04
Non-Hispanic Asian	17.8 (9.4-26.1)	0.7 (0.4-1.3)	.31	0.7 (0.4-1.1)	.10
Non-Hispanic multiple races and other races	17.1 (1.7-32.6)	0.6 (0.2-1.5)	.24	0.6 (0.3-1.6)	.35
Poverty income ratio					
Below poverty level	20.7 (16.1-25.3)	1.0 (Ref)	N/A	1.0 (Ref)	N/A
Between 100% and 200% of poverty level	22.4 (18.4-26.5)	1.0 (0.8-1.4)	.89	1.1 (0.8-1.4	.56
>200% above poverty level	27.4 (24.4-30.5)	*1.8 (1.4-2.3)*	<.001	*1.3 (1.0-1.7)*	.03
Employment status					
Not employed	23.6 (21.3-25.9)	1.0 (Ref)	N/A	1.0 (Ref)	N/A
Employed	33.1 (27.6-38.6)	*2.5 (2.1-3.0)*	<.001	*1.4 (1.2-1.7)*	<.001
Marital status					
Not married	21.4 (18.8-24.1)	1.0 (Ref)	N/A	1.0 (Ref)	N/A
Married	29.8 (26.4-33.3)	*1.6 (1.4-1.9)*	<.001	*1.4 (1.2-1.6)*	<.001
Health status					
Better	24.3 (20.6-28.0)	1.0 (Ref)	N/A	1.0 (Ref)	N/A
Worse	26.3 (22.1-30.5)	0.8 (0.6-1.0)	.07	1.1 (0.9-1.3)	.49
About the same	24.6 (21.6-27.7)	0.8 (0.7-1.0)	.07	1.0 (0.8-1.2)	.90
Education level					
≤High school	12.0 (9.8-14.2)	1.0 (Ref)	N/A	1.0 (Ref)	N/A
Some college	34.9 (30.7-39.2)	*3.3 (2.6-4.0)*	<.001	*2.9 (2.3-3.6)*	<.001
≥Bachelor's degree	44.2 (38.5-50.0)	*4.4 (3.5-5.4)*	<.001	*3.7 (2.9-4.6)*	<.001
Insurance coverage					
Not covered	24.9 (22.7-27.1)	1.0 (Ref)	N/A	1.0 (Ref)	N/A
Covered	26.9 (14.8-36.0)	*1.6 (1.0-2.4)*	.04	1.1 (0.7-1.7)	.72

[a] Italicized values are significant at $P<.05$.

[b] HIT: health information technology. HIT use is a composite variable coded using consecutive "OR" statements for the 5 indicators of HIT use measured in the National Health Interview Survey.

[c] Prevalence ratio adjusted for other sociodemographic characteristics.

[d] Ref: Reference.

[e] N/A: not applicable.

Discussion

In this nationally representative sample of US adults with and without ASCVD, adoption of HIT among adults with a history of ASCVD was lower than that of HIT among adults without a history of ASCVD. HIT use prevalence was also lower among adults with sociodemographic characteristics that indicate vulnerability, such as unemployment and lower levels of education. Previous studies have found that older adults tend to be less likely to use the internet or search for health information online than younger adults [8,20]. Similar trends in HIT use in the older adult population were found in this analysis as the age groups of 60-69 years and 70+ years exhibited the lowest prevalence of HIT use among all age groups. There are several potential barriers to HIT use in the older adult population, such as limited health literacy, poor usability and accessibility of HIT, and impediments to effective use of HIT due to complications from chronic diseases such as vision impairment [21-23]. Care management and health promotion in older adults have been shown to be enhanced by HIT; thus, designing strategies that address these barriers could further enhance its effectiveness [24]. Younger family members and caregivers could be engaged to assist in HIT use.

Vulnerable populations such as those who earn lower incomes and are unemployed are more likely to suffer the consequences of the digital divide. People who are employed and have higher incomes are more likely to communicate with health care providers via text, phone apps, or social media [16]. This finding was corroborated in our analysis as the respondents that were >200% above the poverty level (poverty income ratio>2) and employed had a higher prevalence of HIT use compared to respondents that were below the poverty level and unemployed. Affordability of HIT is thus of major importance to increase access to potentially beneficial interventions that are currently barred by financial constraints. Education level has been cited as a determinant of HIT use in previous literature, with individuals with a college education being more likely to engage in eHealth behaviors compared to individuals without a college education [25]. Lower educational attainment has also been associated with lower health literacy, and this has resulted in individuals seeking self-management support either in person or by phone rather than through the internet [26]. Our analysis was consistent with these findings as respondents with lower educational attainment (a high school education or less) had the lowest prevalence of HIT use. HIT strategies should address gaps linked to health literacy as patients with sufficient health literacy are more likely to have access to the internet at home, search the web, access health information via the internet, email via the internet, and communicate with health care providers than patients with marginal to low health literacy [27].

Evaluating HIT use by race/ethnicity has been a major area of interest in terms of bridging the gap of the digital divide. Previous studies have reported that in both outpatient and inpatient environments, Black and Hispanic people are less likely to adopt and use patient portals [28,29]. A nationally representative study examining the demographics of users of health-related information obtained via the internet found that users were more likely to be White or Asian people [16]. Our analysis of the US adult population with ASCVD also found that there was a statistically significant lower prevalence of HIT use among Non-Hispanic Black people (prevalence ratio of 0.8, P=.04) but not Hispanic people (prevalence ratio of 0.9, P=.43) compared to White people (reference with prevalence ratio of 1.0) when adjusted for other sociodemographic characteristics. However, HIT use in this analysis did not include mobile devices. In recent years, mobile phones have emerged as a method to bridge the digital divide as they can provide access to internet services. Blacks are more likely to use a mobile phone to search for health information via the internet, and these devices could be used for targeted interventions [8].

An analysis using data from the 2012 and 2014 Health Information National Trends Survey measured preferences and use of HIT among US adults with and without 3 chronic disease conditions (CVD, diabetes, and hypertension) [7]. After adjusting for sociodemographic characteristics, the analysis found no significant association between these cardiovascular comorbidities and HIT use, suggesting that sociodemographic factors may have a greater influence on the adoption of HIT than the chronic diseases themselves [7]. Our analysis among US adults with a history of ASCVD that adjusted for sociodemographic variables did find significant differences in the prevalence of overall HIT use. The difference in the overall use of HIT further highlights that our fundamental approach to launching these technologies should account for the socioeconomic challenges they may face. This study demonstrates the influence that sociodemographic characteristics have on HIT adoption. It is important for clinical and public health professionals to incorporate the social determinants that impact patients' health in the design of novel HIT to facilitate effective use. Emerging technologies for patients with ASCVD, such as mobile health interventions, telemedicine, and artificial intelligence have the potential to help manage CVD risk factors, reduce rehospitalizations from cardiac causes, and lower overall health care costs [5,30]. Addressing sociodemographic barriers to HIT use in a population with ASCVD can help ensure that these digital interventions meet the needs of these patients at scale.

This study has the following strengths. The NHIS is a large nationally representative survey, and this analysis pooled 8 years of data to increase the statistical power of the models for more accurate comparisons among the study population.

However, health conditions such as coronary heart disease and stroke were self-reported, which could result in an underestimation of HIT use among this population if some individuals misclassified their condition. The smaller sample sizes for adults within our selection criterion in the race/ethnicity group, Non-Hispanic multiple races and other races, and the age group of 18-29 years may result in our estimates being less precise for those groups. CVD risk factors such as diabetes mellitus, hyperlipidemia, smoking, and hypertension could also serve as additional indicators of ASCVD outcomes and be included in future models assessing HIT use among the ASCVD population.

This study has the following weaknesses. HIT use was defined as using a computer in the past 12 months to perform a health-related task, and this did not include other mobile devices. According to the analysis of the 2018 Health Information National Trends Survey by Shan et al [17], 92% of US adults with or at risk of CVD owned a cell phone and 81% owned a smartphone in 2019. Mobile health devices have the potential to consolidate these services for patient-clinician communication, and their absence in the assessment may have resulted in an underestimation of use.

In conclusion, overall HIT use was low (25%) among adults with self-reported ASCVD, which may represent a barrier to delivering care via emerging HIT. Adults with ASCVD who were older, less educated, unemployed, racial minorities, not married, and had lesser income showed a lower prevalence of overall HIT use. To scale HIT interventions such as telemedicine, targeted strategies are needed to address the sociodemographic barriers to HIT adoption.

Acknowledgments

EMS has received the following financial support for the research, authorship, and publication of this paper: NIH/NHLBI T32 HL007024 Postdoctoral Fellowship in Cardiovascular Epidemiology Institutional Training, NIH/NINR F31 NR017328 Ruth L. Kirschstein National Research Service Award, and NIH/NINR T32 NR012704 Predoctoral Fellowship in Interdisciplinary Cardiovascular Health Research. RATO is supported by the American Heart Association Strategically Focused Research Network on Obesity Award: 17SFRN33590069. YCM and SSM were funded by the American Heart Association Strategically Focused Research Network on Health Technologies and Innovation grant 20SFRN35380046.

Authors' Contributions

UN, RATO, and YCM developed the research concept, acquired the data, and conducted all statistical analyses. UN, RATO, YCM, SSM, TBP, EMS, RS, BK, and DB contributed to writing and reviewing the manuscript. UN, RATO, EMS, and YCM provided statistical expertise. RATO and YCM provided supervision. All authors reviewed the final manuscript.

Conflicts of Interest

Under a license agreement between Corrie Health and the Johns Hopkins University, the University owns equity in Corrie Health and the University and SSM are entitled to royalty distributions. Additionally, SSM is a co-founder of and holds equity in Corrie Health. This arrangement has been reviewed and approved by the Johns Hopkins University in accordance with its conflict of interest policies. In addition, SSM has served as a scientific advisor to Amgen, AstraZeneca, Dalcor Pharmaceuticals, Esperion, iHealth, Kaneka, Novartis, Regeneron, REGENXBIO, Sanofi, and 89bio.

References

1. Benjamin EJ, Muntner P, Alonso A, Bittencourt MS, Callaway CW, Carson AP, American Heart Association Council on EpidemiologyPrevention Statistics CommitteeStroke Statistics Subcommittee. Heart Disease and Stroke Statistics-2019 Update: A Report From the American Heart Association. Circulation 2019 Mar 05;139(10):e56-e528. [doi: 10.1161/CIR.0000000000000659] [Medline: 30700139]
2. Karunathilake SP, Ganegoda GU. Secondary Prevention of Cardiovascular Diseases and Application of Technology for Early Diagnosis. Biomed Res Int 2018;2018:5767864 [FREE Full text] [doi: 10.1155/2018/5767864] [Medline: 29854766]
3. Havranek EP, Mujahid MS, Barr DA, Blair IV, Cohen MS, Cruz-Flores S, American Heart Association Council on Quality of CareOutcomes Research, Council on EpidemiologyPrevention, Council on CardiovascularStroke Nursing, Council on LifestyleCardiometabolic Health,Stroke Council. Social Determinants of Risk and Outcomes for Cardiovascular Disease: A Scientific Statement From the American Heart Association. Circulation 2015 Sep 01;132(9):873-898. [doi: 10.1161/CIR.0000000000000228] [Medline: 26240271]
4. Vollmer WM, Owen-Smith AA, Tom JO, Laws R, Ditmer DG, Smith DH, et al. Improving adherence to cardiovascular disease medications with information technology. Am J Manag Care 2014 Nov;20(11 Spec No. 17):SP502-SP510 [FREE Full text] [Medline: 25811824]
5. Molinari G, Molinari M, Di Biase M, Brunetti ND. Telecardiology and its settings of application: An update. J Telemed Telecare 2018 Jun;24(5):373-381. [doi: 10.1177/1357633X16689432] [Medline: 28084886]
6. Alharbi NS, Alsubki N, Jones S, Khunti K, Munro N, de Lusignan S. Impact of Information Technology-Based Interventions for Type 2 Diabetes Mellitus on Glycemic Control: A Systematic Review and Meta-Analysis. J Med Internet Res 2016 Nov 25;18(11):e310 [FREE Full text] [doi: 10.2196/jmir.5778] [Medline: 27888169]

7. Asan O, Cooper Ii F, Nagavally S, Walker RJ, Williams JS, Ozieh MN, et al. Preferences for Health Information Technologies Among US Adults: Analysis of the Health Information National Trends Survey. J Med Internet Res 2018 Oct 18;20(10):e277 [FREE Full text] [doi: 10.2196/jmir.9436] [Medline: 30341048]
8. Fox S, Duggan M. Health online 2013. Pew Research Center. URL: https://www.pewinternet.org/wp-content/uploads/sites/9/media/Files/Reports/PIP_HealthOnline.pdf [accessed 2020-05-20]
9. 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]
10. Coronavirus disease (COVID-19) pandemic. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 [accessed 2021-01-16]
11. Mehrotra A, Chernew M. As outpatient visits rebound, COVID-19 pandemic's threat to outpatient care remains. The Commonwealth Fund. URL: https://www.commonwealthfund.org/publications/2020/apr/impact-covid-19-outpatient-visits [accessed 2021-06-29]
12. Huh J, Koola J, Contreras A, Castillo AK, Ruiz M, Tedone KG, et al. Consumer Health Informatics Adoption among Underserved Populations: Thinking beyond the Digital Divide. Yearb Med Inform 2018 Aug;27(1):146-155 [FREE Full text] [doi: 10.1055/s-0038-1641217] [Medline: 30157518]
13. Anderson M. Mobile technology and home broadband 2019. Pew Research Center. URL: https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2019/06/PI_2019.06.13_Mobile-Technology-and-Home-Broadband_FINAL2.pdf [accessed 2020-05-16]
14. Atkinson NL, Saperstein SL, Pleis J. Using the internet for health-related activities: findings from a national probability sample. J Med Internet Res 2009 Feb 20;11(1):e4 [FREE Full text] [doi: 10.2196/jmir.1035] [Medline: 19275980]
15. 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]
16. Wimble M. Understanding Health and Health-Related Behavior of Users of Internet Health Information. Telemed J E Health 2016 Oct;22(10):809-815. [doi: 10.1089/tmj.2015.0267] [Medline: 27045569]
17. Shan R, Ding J, Plante TB, Martin SS. Mobile Health Access and Use Among Individuals With or At Risk for Cardiovascular Disease: 2018 Health Information National Trends Survey (HINTS). J Am Heart Assoc 2019 Dec 17;8(24):e014390 [FREE Full text] [doi: 10.1161/JAHA.119.014390] [Medline: 31818220]
18. National Health Interview Survey, 2011-2018. Public-use data file and documentation. National Center for Health Statistics. URL: https://www.cdc.gov/nchs/nhis/1997-2018.htm [accessed 2019-08-03]
19. National Health Interview Survey - Methods. National Center for Health Statistics. URL: https://www.cdc.gov/nchs/nhis/methods.htm [accessed 2019-08-03]
20. Fox S, Rainie L. The web at 25 in the US. Pew Research Center. URL: https://www.pewresearch.org/internet/2014/02/27/the-web-at-25-in-the-u-s/ [accessed 2020-05-16]
21. Martinez VI, Marquard JL, Saver B, Garber L, Preusse P. Consumer Health Informatics Interventions Must Support User Workflows, Be Easy-To-Use, and Improve Cognition: Applying the SEIPS 2.0 Model to Evaluate Patients' and Clinicians' Experiences with the CONDUIT-HID Intervention. International Journal of Human–Computer Interaction 2017 Jan 11;33(4):333-343. [doi: 10.1080/10447318.2016.1278340]
22. Wildenbos GA, Peute L, Jaspers M. Facilitators and Barriers of Electronic Health Record Patient Portal Adoption by Older Adults: A Literature Study. Stud Health Technol Inform 2017;235:308-312. [Medline: 28423804]
23. Isaković M, Sedlar U, Volk M, Bešter J. Usability Pitfalls of Diabetes mHealth Apps for the Elderly. J Diabetes Res 2016;2016:1604609 [FREE Full text] [doi: 10.1155/2016/1604609] [Medline: 27034957]
24. Chiu C, Wu CH. Information and Communications Technology as a Health Promotion Method for Older Adults in Assisted-Living Facilities: Three-Arm Group-Randomized Trial. JMIR Aging 2019 May 06;2(1):e12633 [FREE Full text] [doi: 10.2196/12633] [Medline: 31518261]
25. Amo L. Education-Based Gaps in eHealth: A Weighted Logistic Regression Approach. J Med Internet Res 2016 Oct 12;18(10):e267 [FREE Full text] [doi: 10.2196/jmir.5188] [Medline: 27733329]
26. Sarkar U, Piette JD, Gonzales R, Lessler D, Chew LD, Reilly B, et al. Preferences for self-management support: findings from a survey of diabetes patients in safety-net health systems. Patient Educ Couns 2008 Jan;70(1):102-110 [FREE Full text] [doi: 10.1016/j.pec.2007.09.008] [Medline: 17997264]
27. 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]
28. Yamin CK, Emani S, Williams DH, Lipsitz SR, Karson AS, Wald JS, et al. The digital divide in adoption and use of a personal health record. Arch Intern Med 2011 Mar 28;171(6):568-574. [doi: 10.1001/archinternmed.2011.34] [Medline: 21444847]
29. Elston Lafata J, Miller CA, Shires DA, Dyer K, Ratliff SM, Schreiber M. Patients' adoption of and feature access within electronic patient portals. Am J Manag Care 2018 Nov 01;24(11):e352-e357 [FREE Full text] [Medline: 30452203]

30. Widmer RJ, Allison TG, Lerman LO, Lerman A. Digital Health Intervention as an Adjunct to Cardiac Rehabilitation Reduces Cardiovascular Risk Factors and Rehospitalizations. J Cardiovasc Transl Res 2015 Jul;8(5):283-292 [FREE Full text] [doi: 10.1007/s12265-015-9629-1] [Medline: 25946990]

Abbreviations

ASCVD: atherosclerotic cardiovascular disease
HIT: health information technology
NCHS: National Center for Health Statistics
NHIS: National Health Interview Survey

Edited by G Eysenbach; submitted 22.08.20; peer-reviewed by J Marquard, S Speedie, S Nagavally; comments to author 14.01.21; revised version received 07.02.21; accepted 06.04.21; published 13.08.21.

Please cite as:
Nwokeji U, Spaulding EM, Shan R, Turkson-Ocran RA, Baptiste D, Koirala B, Plante TB, Martin SS, Commodore-Mensah Y
Health Information Technology Use Among Persons With Self-reported Atherosclerotic Cardiovascular Disease: Analysis of the 2011-2018 National Health Interview Survey
J Med Internet Res 2021;23(8):e23765
URL: https://www.jmir.org/2021/8/e23765
doi:10.2196/23765
PMID:34397391

©Uchenna Nwokeji, Erin M Spaulding, Rongzi Shan, Ruth-Alma Turkson-Ocran, Diana Baptiste, Binu Koirala, Timothy B Plante, Seth S Martin, Yvonne Commodore-Mensah. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Obesity and BMI Cut Points for Associated Comorbidities: Electronic Health Record Study

Natalie Liu[1], MD; Jen Birstler[2], MSc; Manasa Venkatesh[1], MSc; Lawrence Hanrahan[3], MSc, PhD; Guanhua Chen[2], PhD; Luke Funk[1,4], MD, MPH

[1]Department of Surgery, University of Wisconsin-Madison, Madison, WI, United States
[2]Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI, United States
[3]Department of Family Medicine and Community Health, University of Wisconsin-Madison, Madison, WI, United States
[4]Department of Surgery, William S. Middleton Memorial VA, Madison, WI, United States

Corresponding Author:
Luke Funk, MD, MPH
Department of Surgery
University of Wisconsin-Madison
600 Highland Ave
Madison, WI, 53792
United States
Phone: 1 6082631036
Fax: 1 6082520942
Email: funk@surgery.wisc.edu

Abstract

Background: Studies have found associations between increasing BMIs and the development of various chronic health conditions. The BMI cut points, or thresholds beyond which comorbidity incidence can be accurately detected, are unknown.

Objective: The aim of this study is to identify whether BMI cut points exist for 11 obesity-related comorbidities.

Methods: US adults aged 18-75 years who had ≥3 health care visits at an academic medical center from 2008 to 2016 were identified from eHealth records. Pregnant patients, patients with cancer, and patients who had undergone bariatric surgery were excluded. Quantile regression, with BMI as the outcome, was used to evaluate the associations between BMI and disease incidence. A comorbidity was determined to have a cut point if the area under the receiver operating curve was >0.6. The cut point was defined as the BMI value that maximized the Youden index.

Results: We included 243,332 patients in the study cohort. The mean age and BMI were 46.8 (SD 15.3) years and 29.1 kg/m^2, respectively. We found statistically significant associations between increasing BMIs and the incidence of all comorbidities except anxiety and cerebrovascular disease. Cut points were identified for hyperlipidemia (27.1 kg/m^2), coronary artery disease (27.7 kg/m^2), hypertension (28.4 kg/m^2), osteoarthritis (28.7 kg/m^2), obstructive sleep apnea (30.1 kg/m^2), and type 2 diabetes (30.9 kg/m^2).

Conclusions: The BMI cut points that accurately predicted the risks of developing 6 obesity-related comorbidities occurred when patients were overweight or barely met the criteria for class 1 obesity. Further studies using national, longitudinal data are needed to determine whether screening guidelines for appropriate comorbidities may need to be revised.

(J Med Internet Res 2021;23(8):e24017) doi:10.2196/24017

KEYWORDS

obesity; body mass index (BMI); risk factors; screening; health services; chronic disease

Introduction

Background

Obesity (BMI ≥30.0 kg/m^2) is a global public health problem. The highest rates of obesity occur in the United States, where over one-third of adults have obesity [1]. In 1998, the World Health Organization created international standardized BMI classifications for adults who are overweight and have obesity based on risks of obesity-related diseases for European adults [2]. These classifications were based on the risks of obesity-related diseases in European adults with varied BMI values [3]. On the basis of these classifications, overweight and obesity were defined as having a BMI between 25.0 and 29.9 kg/m^2 and a BMI≥30.0 kg/m^2, respectively. However, studies have demonstrated that the risks of obesity-related comorbidities differ based on sex and race or ethnicity. Female Asian patients have been shown to develop comorbidities at lower BMIs, suggesting that BMI thresholds for overweight and obesity should be lower for these groups [2,4-7].

Study Significance

Obesity is associated with numerous comorbidities, including hypertension, hyperlipidemia, type 2 diabetes mellitus (T2DM), and coronary artery disease (CAD) [8-10]. The cross-sectional study by Pantalone et al [8], which used electronic health record (EHR) data, showed that patients with higher BMIs had a higher prevalence of T2DM, hypertension, and CAD. However, studies have not addressed whether specific BMI cut points exist for US adults. BMI cut points are defined as the thresholds beyond which disease incidence can be accurately detected. In addition, no studies have evaluated cut points by using EHR data that provide patient-level information for large, multiethnic cohorts. Studies have concluded that it is feasible to use EHR analysis to study chronic diseases such as obesity, diabetes, and hypertension [11,12].

Objective

The objective of this study is to examine EHR data from a large health care system in the United States to determine whether BMI cut points exist for 11 common comorbidities associated with obesity and being overweight. We also evaluate whether cut points varied with sex and race or ethnicity. We hypothesize that most cut points would occur in the class 1 obesity category.

Methods

Data Source

We used data from the University of Wisconsin Hospital and Clinics EHR over a 10-year period (June 1, 2008, to December 31, 2018). All patient data and analyses were stored on a secure server managed through the University of Wisconsin Health Information Services and the Institute for Clinical and Translational Research. The Epic Clarity Database was used as the data source for all patients. This study was approved by the University of Wisconsin Minimal Risk institutional review board (protocol #2017-0443), and the need for informed consent was waived. We followed the Strengthening the Reporting of Observational Studies in Epidemiology guidelines within the Enhancing the Quality and Transparency of Health Research network in the methodology and reporting of this study (Multimedia Appendix 1 contains the full Strengthening the Reporting of Observational Studies in Epidemiology checklist) [13].

Data Validation and Cleaning

All recorded heights and weights in the EHR were cleaned to reduce the inclusion of incorrect heights and weights because of errors in data entry. Similar to our previous study using EHR data, we used the methodology proposed by Cheng et al [14] to remove biologically implausible heights and weights [15]. All heights >90 inches, <44 inches, and >1 SD from the mean height when SD was >2.5% of the mean were removed. All weights >1000 pounds, <55 pounds, >70% of the range from the mean when the range ≥50 pounds, and >1 SD from the mean when the SD was >20% of the mean were removed. Missing height data were imputed with the most recent previous nonmissing valid height. Any remaining missing height was replaced with the most recent subsequent nonmissing valid height. BMI values were calculated using the valid heights and weights. No patients were excluded from the study because of the data cleaning process.

Study Population

We included all patients between the ages of 18 and 75 years who had ≥3 in-person clinical visits over a minimum of 2 years documented in the EHR during the study period. All included patients had an *index visit* with a valid BMI measurement, another visit at least 1 year before the index visit, and an additional visit 1 year after their index visit. The minimum 1-year period between the index visit and the previous visit was used to identify patients who had each disease of interest versus those who did not. The 1-year period between the index visit and the subsequent visit was used to calculate 1-year incidence rates for patients who did not have the disease before the index visit but were later diagnosed with the disease. Patients with multiple intervals of ≥3 clinical visits had an interval selected at random.

Patients with a pregnancy or cancer diagnosis at any time before or during the study period were excluded using the International Classification of Disease (ICD)-9 and ICD-10 codes. Patients who had undergone bariatric surgery were identified from our institutional bariatric surgery registry and excluded.

Study Variables

Baseline BMI (BMI at the index visit), age (at the index visit), sex (male or female), race or ethnicity (White, non-Hispanic; Black, non-Hispanic; Asian, non-Hispanic; Native American, non-Hispanic; Hispanic; or other or unspecified), insurance type (commercial or private insurance, Medicare, Medicaid, or other or unspecified), and smoking status (at the index visit; active smoker, former smoker, passive smoker [defined as an individual who has had exposure to tobacco smoke but has never smoked themselves], or nonsmoker) were identified from the EHR. Insurance type was defined as the insurance type used during or before the index visit.

Through a literature review, we identified 11 common obesity-related comorbidities that were included in this study: anxiety, CAD, cerebrovascular disease, chronic pain, depression, gastroesophageal reflux disease, hyperlipidemia, hypertension, obstructive sleep apnea (OSA), osteoarthritis, and T2DM [8,9,16,17]. Incident cases were defined as patients who did not have the disease before the index visit and subsequently developed the disease after the index visit. The 1-year incidence rates (defined per 100 person-years) were calculated based on the occurrence of an ICD-9 or ICD-10 code (Multimedia Appendix 2 contains the full list of ICD-9 and ICD-10 codes) during the 1-year period following the index visit for patients who did not have a diagnosis before the index visit. Prevalent cases were defined as patients who had a diagnosis of comorbidity at or before the index visit and identified using the occurrence of an ICD-9 or ICD-10 code during this time.

Statistical Analysis

We used quantile regression with BMI as the outcome to identify differences in the median BMIs between incident cases of each comorbidity and those who did not develop each comorbidity. Two models were fit for each comorbidity to evaluate the associations between BMI and disease incidence—an unadjusted model with disease incidence as the only independent variable and an adjusted model accounting for baseline age, sex, race or ethnicity, and smoking status. We used quantile regression because we were unable to meet the assumptions of the linear model. Quantile regression also allowed for the evaluation of differences in BMI distributions among patients who developed each comorbidity versus those who did not, which is more informative than differences in single mean values [18]. The difference in median BMIs (the median BMI of incident cases minus the median BMI of patients who did not develop the disease) was the outcome of the quantile model.

We conducted cut point analyses with BMI as a screening test for the incidence of each obesity-related comorbidity. Sensitivity and specificity were calculated for continuous BMI values. A comorbidity had a BMI cut point if the area under the receiver operating curve (AUROC) was >0.6. We chose an AUROC>0.6 to ensure that cut points had significant diagnostic value. Although there is no gold standard method, other investigators have used AUROC thresholds that range from >0.5 to >0.7 to determine cut points [6]. For all comorbidities with an AUROC >0.6, the cut point was defined as the BMI value that maximized the Youden index (sensitivity+specificity-1). BMI cut points were also calculated by sex and race or ethnicity and compared using the bootstrap method with 1000 resamplings. The overall incidence rates above and below each cut point were calculated. For any comorbidities that had an identifiable cut point, baseline characteristics and prevalence of any concurrent comorbidities were compared between patients who developed the comorbidity and those who did not develop the comorbidity.

All statistical analyses were conducted using R version 3.6.3 (R Foundation for Statistical Computing).

Incidence Versus Prevalence Cut Point Analysis

Studies have identified cut points for diseases such as diabetes, hypertension, and hyperlipidemia using both incidence and prevalence [6,19]. As there is no standardized method to determine cut points, we analyzed cut point differences between prevalent and incident cases. For any comorbidities that had an identifiable cut point, we used the bootstrap method with 1000 resamplings to determine cut points and *P* values comparing incident and prevalent cases.

Results

Patient Characteristics

Over 300,000 patients had at least three clinical visits during the study period. After applying exclusion criteria, 243,332 patients met inclusion criteria (Figure 1). The mean age was 46.8 (SD 15.3) years (Table 1). Of the patients, 54.9% (133,654/243,332) of the patients were female, and 88.7% (215,950/243,332) patients were White and non-Hispanic. The mean BMI was 29.1 (SD 7.0) kg/m^2, and 36.8% (89,660/243,332) of patients had a BMI ≥30 kg/m^2. In our study cohort, 57.7% (139,753/243,332) of patients had never smoked or used tobacco products, whereas 14.1% (34,328/243,332) of patients were active smokers. Hyperlipidemia and hypertension were the most common comorbidities, affecting 24.3% (59,097/243,332) patients and 21.5% (52,365/243,332) of the study population, respectively (Table 1).

Figure 1. Study cohort creation (Strengthening the Reporting of Observational Studies in Epidemiology diagram). EHR: electronic health record.

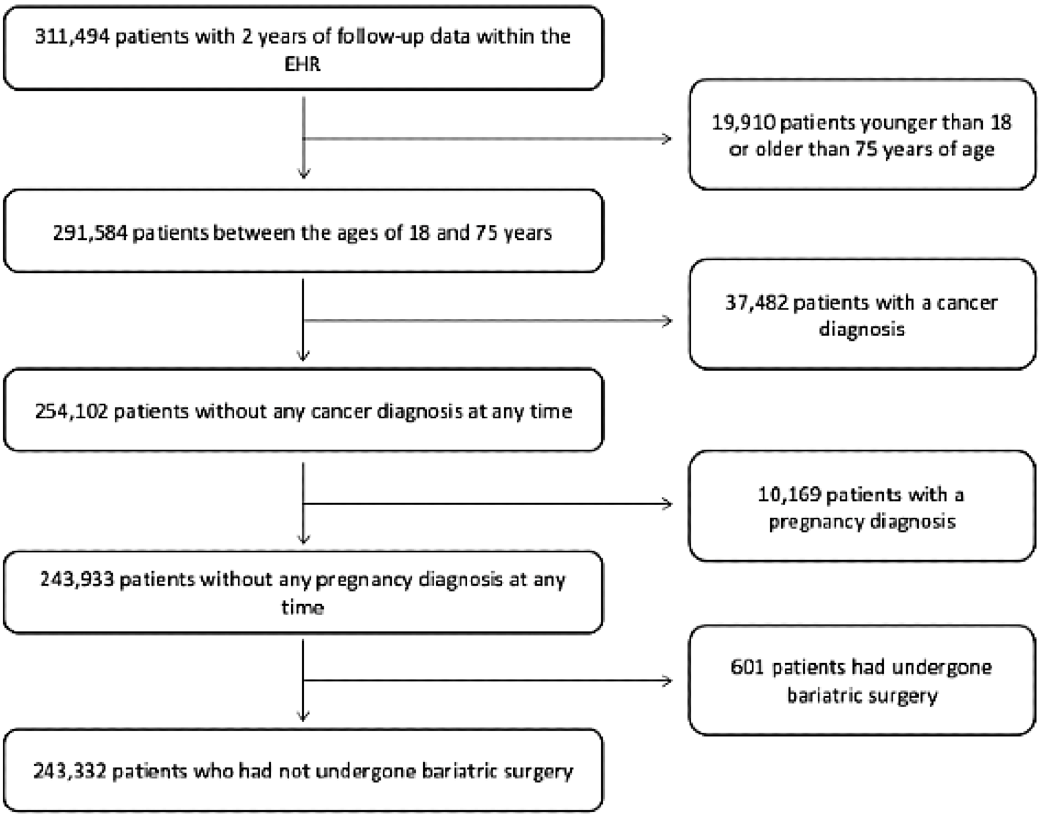

311,494 patients with 2 years of follow-up data within the EHR
19,910 patients younger than 18 or older than 75 years of age
291,584 patients between the ages of 18 and 75 years
37,482 patients with a cancer diagnosis
254,102 patients without any cancer diagnosis at any time
10,169 patients with a pregnancy diagnosis
243,933 patients without any pregnancy diagnosis at any time
601 patients had undergone bariatric surgery
243,332 patients who had not undergone bariatric surgery

Table 1. Baseline demographics and patient characteristics (N=243,332).

Characteristics	Values
Age (years), mean (SD)	46.8 (15.3)
Sex, n (%)	
Male	109,678 (45.1)
Female	133,654 (54.9)
Race or ethnicity, n (%)	
White, non-Hispanic	215,950 (88.7)
Black, non-Hispanic	9463 (3.9)
Asian, non-Hispanic	6621 (2.7)
Native American, non-Hispanic	1161 (0.5)
Hispanic	7730 (3)
Other or unspecified	2767 (1.1)
Baseline BMI category (kg/m^2)[a], n (%)	
Underweight (<18.5)	3000 (1.2)
Normal (18.5-24.9)	72,803 (29.9)
Overweight (25.0-29.9)	77,869 (32)
Class 1 obesity (30.0-34.9)	48,213 (19.8)
Class 2 obesity (35.0-39.9)	23,371 (9.6)
Class 3 obesity (>40)	18,076 (7.4)
Insurance type, n (%)	
Commercial	191,697 (78.8)
Medicare	31,778 (5.7)
Medicaid	6032 (2.5)
Other or unspecified	13,825 (5.7)
Prevalence of comorbidities, n (%)	
Anxiety	33,984 (14)
Coronary artery disease	9543 (3.9)
Cerebrovascular disease	3076 (1.3)
Chronic pain	14,479 (6)
Depression	32,210 (13.2)
Gastroesophageal reflux	29,512 (12.1)
Hyperlipidemia	59,097 (24.3)
Hypertension	52,365 (21.5)
Obstructive sleep apnea	13,746 (5.6)
Osteoarthritis	21,408 (8.8)
Type 2 diabetes mellitus	18,182 (7.5)
Smoking status, n (%)	
Active smoker	34,328 (14.1)
Former smoker	64,331 (26.4)
Passive smoker	2746 (1.1)
Nonsmoker	139,753 (57.4)

[a] The mean baseline BMI was 29.1 kg/m^2 (SD 7.0 kg/m^2).

Incidence of 11 Comorbidities and Their Associations With BMI

The highest 1-year incidence rates were for hyperlipidemia (4.0 cases per 100 person-years) and hypertension (3.6 cases per 100 person-years; Multimedia Appendix 3 contains the full table of 1-year incidence rates). CAD and cerebrovascular disease had the lowest 1-year incidence rates (0.9 and 0.4 cases per 100-person-years, respectively).

In quantile regression, when comparing the median BMI of those who developed each comorbidity (incident group) versus the median BMI of those who did not, we found statistically significant differences in the median BMIs for all obesity-related comorbidities (Multimedia Appendix 4 contains the full table of the quantile regression analysis evaluating associations between BMI and comorbidity incidence). The median BMIs of the incident groups were higher for all comorbidities except for anxiety (–0.6 kg/m^2; 95% CI –0.8 to –0.4).

After adjusting for age, sex, race or ethnicity, and smoking status, we found statistically significant differences in the median BMIs for all comorbidities except anxiety and cerebrovascular disease (Multimedia Appendix 4 contains the full table of the quantile regression analysis evaluating associations between BMI and comorbidity incidence). The adjusted median BMIs of the incident groups were higher for all comorbidities. The greatest differences in adjusted median BMI were for OSA (6.0 kg/m^2; 95% CI 5.7-6.4) and T2DM (5.0 kg/m^2; 95% CI 4.6-5.4).

BMI Cut Points for All Study Patients

Six comorbidities had BMI cut points: CAD, hyperlipidemia, hypertension, OSA, osteoarthritis, and T2DM (Table 2). Hyperlipidemia had the lowest cut point (27.1 kg/m^2; sensitivity=68.8%; specificity=52.1%), followed by CAD (27.7 kg/m^2; sensitivity=66.5%; specificity=50.5%), hypertension (28.4 kg/m^2; sensitivity=62.3%; specificity=60.7%), osteoarthritis (28.7 kg/m^2; sensitivity=58.7%; specificity=51.7%), OSA (30.1 kg/m^2; sensitivity=72%; specificity=66.6%), and T2DM (30.9 kg/m^2; sensitivity=63.3%; specificity=70.9%).

The 1-year incidence rates above the cut point were higher than the rates below the cut point for the six comorbidities that had identified cut points (Figure 2). The greatest differences were for OSA (0.7 cases per 100 person-years below vs 3.4 cases per 100 person-years above the cut point) and T2DM (0.6 cases per 100 person-years below vs 2.5 cases per 100 person-years above the cut point).

When comparing baseline demographics for the comorbidities with an identifiable cut point (CAD, hyperlipidemia, hypertension, OSA, osteoarthritis, and T2DM), we found that patients who developed each disease were older and more likely to be male than those who did not develop each disease for all six comorbidities (Multimedia Appendices 5-10 contain tables comparing baseline characteristics of patients who developed each comorbidity vs those who did not for all six comorbidities with a cut point). Patients who developed each comorbidity had a higher prevalence of each of the other five comorbidities with an identifiable cut point. For example, patients who developed hypertension had a higher prevalence of CAD, hyperlipidemia, OSA, osteoarthritis, and T2DM.

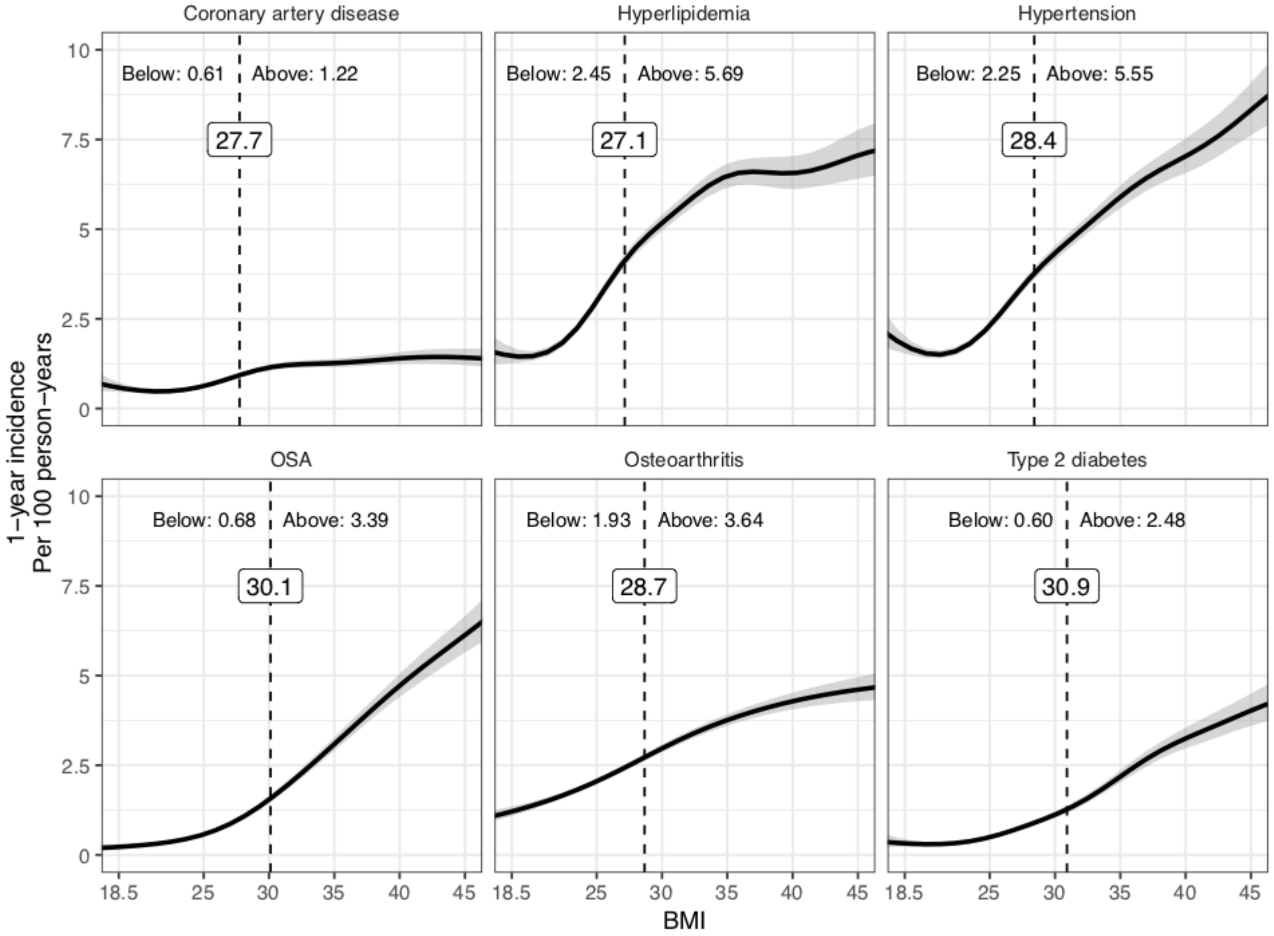
Table 2. Cut points for comorbidities.

Comorbidity	AUROC[a]	Youden index	Sensitivity, %	Specificity, %	Cut point (kg/m^2)
Anxiety	0.477	N/A[b]	N/A	N/A	N/A
Coronary artery disease	0.603	0.170	66.5	50.5	27.7
Cerebrovascular disease	0.561	N/A	N/A	N/A	N/A
Chronic pain	0.559	N/A	N/A	N/A	N/A
Depression	0.521	N/A	N/A	N/A	N/A
Gastroesophageal reflux	0.555	N/A	N/A	N/A	N/A
Hyperlipidemia	0.637	0.209	68.8	52.1	27.1
Hypertension	0.653	0.230	62.3	60.7	28.4
Obstructive sleep apnea	0.754	0.386	72	66.6	30.1
Osteoarthritis	0.606	0.161	58.7	51.7	28.7
Type 2 diabetes mellitus	0.725	0.341	63.3	70.9	30.9

[a] AUROC: area under the receiver operating curve.

[b] N/A: not applicable.

Figure 2. Cut points and comorbidity incidence. Gray shaded areas represent 95% CIs. The dotted line and the values in the box represent BMI cut points. "Below" corresponds to overall disease incidence (per 100 person-years) for all patients with a BMI that is less than the cut point. "Above" corresponds to overall disease incidence (per 100 person-years) for all patients with a BMI that is greater than the cut point. OSA: obstructive sleep apnea.



BMI Cut Points by Sex

Both male and female patients had cut points for hyperlipidemia, hypertension, OSA, and T2DM, but only female patients had cut points for CAD and osteoarthritis (Table 3; Multimedia Appendix 11 contains the full table of AUROC, Youden index, sensitivity, and specificity values for cut points by sex and race or ethnicity). Female patients had a statistically significant lower cut point for T2DM (29.9 vs 32.1 kg/m^2; P=.02). There were no differences in other cut points between the male and female patients.

Table 3. Cut points by sex.

Comorbidity	Male cut point (kg/m^2)	Female cut point (kg/m^2)	P value[a]
Coronary artery disease	N/A[b]	27.8	N/A
Hyperlipidemia	28.3	28.6	.78
Hypertension	28.8	28.5	.84
Obstructive sleep apnea	31.3	30.2	.74
Osteoarthritis	N/A	29.2	N/A
Type 2 diabetes mellitus	32.1	29.9	.02

[a] P value indicates the comparison of cut points between male and female patients.

[b] N/A: not applicable.

BMI Cut Points by Race or Ethnicity

When evaluating cut points by race or ethnicity, Black patients had higher cut points for hypertension (30.3 vs 28.7 kg/m^2 for White patients; P<.001) and OSA (35.1 vs 30.1 kg/m^2; P=.005; Table 4; Multimedia Appendix 11 contains the full table of AUROC, Youden index, sensitivity, and specificity values for cut points by sex and race or ethnicity). Asian patients had lower cut points for hyperlipidemia (24.1 vs 26.5 kg/m^2 for White patients; P=.02), OSA (29.0 vs 30.1 kg/m^2; P=.02), and T2DM (27.5 vs 31.3 kg/m^2; P=.04). Native American patients had lower cut points for hypertension (26.0 vs 28.7 kg/m^2 for White patients) and T2DM (29.3 vs 31.3 kg/m^2) and a higher cut point for hyperlipidemia (28.8 vs 26.5 kg/m^2), but these differences were not statistically significant. For Hispanic patients, we only identified a cut point for OSA (31.3 kg/m^2; sensitivity=69.2%; specificity=70.4%).

Table 4. Cut points by race or ethnicity.

Comorbidity	White, non-Hispanic	Black, non-Hispanic		Asian, non-Hispanic		Native American, non-Hispanic		Hispanic	
	Cut point (kg/m^2)	Cut point (kg/m^2)	*P* value[a]	Cut point (kg/m^2)	*P* value	Cut point (kg/m^2)	*P* value	Cut point (kg/m^2)	*P* value
Coronary artery disease	27.4	N/A[b]	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Hyperlipidemia	26.5	N/A	N/A	24.1	.02	28.8	.41	N/A	N/A
Hypertension	28.7	30.3	.001	25.0	.23	26.0	.45	N/A	N/A
Obstructive sleep apnea	30.1	35.1	.005	29.0	.02	N/A	N/A	31.3	.08
Osteoarthritis	28.7	31.0	.55	N/A	N/A	N/A	N/A	N/A	N/A
Type 2 diabetes mellitus	31.3	31.3	.91	27.5	.04	29.3	.55	N/A	N/A

[a] *P* value indicates the comparison to cut points for White, non-Hispanic patients.

[b] N/A: not applicable.

Incidence Versus Prevalence Cut Point Analysis

For the six comorbidities that had BMI cut points, we found no statistically significant differences in cut points between the incident and prevalent cases for CAD, hypertension, OSA, and osteoarthritis (Multimedia Appendix 12 contains the full table of incidence vs prevalence cut points). There were statistically significant differences between incidence and prevalence cut points for hyperlipidemia (27.5 vs 27.0 kg/m^2; P=.02) and T2DM (30.7 vs 30.0 kg/m^2; P<.001).

Discussion

Principal Findings

Our findings suggest that the BMI cut points or thresholds beyond which disease incidence can be accurately detected for developing six obesity-related comorbidities occur when patients are overweight or barely meet the criteria for class 1 obesity. The cut points for developing CAD, hyperlipidemia, hypertension, and osteoarthritis were in the overweight category, while the cut points for OSA and T2DM occurred at the transition between overweight and class 1 obesity. In our study cohort, female patients had lower cut points for T2DM. Asian patients had lower cut points for hyperlipidemia, OSA, and T2DM, while Black patients had higher cut points for hypertension and OSA.

Most cut points identified in our study were within the overweight BMI category. Published studies are currently mixed with regard to the association between being overweight and the development of *obesity-related* comorbidities. The meta-analysis by Guh et al [9] found that the relative risks for comorbidities, such as T2DM and CAD, increased when patients were overweight but increased most when patients were obese. Other studies, such as the cross-sectional study by Nguyen et al [16], which used National Health and Nutrition Examination Survey (NHANES) data, demonstrated that higher BMIs were associated with an increased risk of these diseases. In contrast, a retrospective cohort study of Swiss adults by Faeh et al [10] showed increased mortality rates in patients with obesity because of CAD but not in patients who were overweight. Despite numerous studies identifying associations between these chronic diseases and obesity, no studies have identified these cut points in multiracial or ethnic populations.

We found that female patients had a lower cut point for T2DM than male patients. The literature is inconclusive regarding the association between sex and the development of obesity-related comorbidities. The retrospective study by Chu et al [6] found lower cut points for both hypertension and T2DM in Taiwanese women than men. A large cohort study evaluating the incidence of hypertension in Japanese adults with obesity showed that the relationship between BMI and hypertension was influenced by sex, with female patients experiencing a greater risk of developing hypertension [20]. In contrast, a retrospective study by Ong et al [21] of US adults using data from NHANES showed no difference in the risk of hypertension between men and women. Although our results showed no differences in hyperlipidemia cut points between male and female patients, a retrospective cohort study by Tseng et al [19] demonstrated a lower cut point for hyperlipidemia in Taiwanese women than men.

Our study found that compared with White patients, Black patients had higher cut points for hypertension and OSA. The cross-sectional study by Fontaine et al [22] using NHANES data found that Black patients experienced obesity-related morbidity, such as reduction in lifespan, at higher BMIs than White patients. In a review, Wagner and Heyward [23] hypothesized that differences in the development of obesity-related comorbidities between Black patients and those of different racial or ethnic backgrounds stemmed from variations in body composition; Black patients typically have

XSL•FO RenderX

higher BMIs than White patients despite having similar levels of body fat.

We also found that Asian patients had lower cut points. This is supported by the Expert Committee of the World Health Organization, which concluded that Asian populations have different associations between BMI and obesity-related diseases and that the cut points of obesity-related comorbidities in Asians varied between 22.0-25.0 kg/m^2 [4,7]. The population-based cross-sectional study by Cheong et al [24] of Malaysian adults identified BMI cut points for predicting the presence of diabetes, hypertension, and hyperlipidemia to be between 23.3-24.1 kg/m^2 in men and 24.0-25.4 kg/m^2 in women. A prospective study by Chan et al [25] of Chinese adults diagnosed with CAD identified a BMI cut point of 27.3 kg/m^2 for the development of OSA. The lower cut points in Asian patients have been attributed to a multitude of genetic and metabolic differences between Asian and White patients, such as different associations between BMI and body fat percentage in Asian versus White populations [4,7]. In addition, there may be differences among the various Asian subgroups. A secondary analysis by Jih et al [7] of the California Health Interview Survey found the highest rates of overweight or obesity and diabetes in Filipino populations, suggesting that genetic, lifestyle, and dietary factors may account for variations in cut points and disease risk.

Study Implications

Our results suggest that although some current screening guidelines incorporating BMI have appropriate cut points, others may need to be revised. For example, the United States Preventative Services Task Force (USPSTF) recommends screening for T2DM [26] and hypertension [27] at a BMI cut point of 25 kg/m^2. Our BMI cut points of 30.9 kg/m^2 and 28.4 kg/m^2 for T2DM and hypertension, respectively, support these guideline cut points.

In contrast, guidelines for OSA screening vary. The American Academy of Sleep Medicine recommends OSA screening for adults with a BMI ≥30 kg/m^2 [28]. The American Federal Aviation Administration and the US Federal Motor Carrier Safety Administration suggest that pilots with BMI ≥40 kg/m^2 and drivers with BMI ≥35 kg/m^2, respectively, should be screened for OSA [29,30]. The American Academy of Sleep Medicine BMI cut point of 30 kg/m^2 and our cut point of 30.1 kg/m^2 suggest that the Federal Aviation Administration and US Federal Motor Carrier Safety Administration screening cut points for OSA may be too high.

BMI is not included in the current screening recommendations for hyperlipidemia, CAD, or osteoarthritis. Although the USPSTF and American College of Cardiology/American Heart Association have guidelines for hyperlipidemia screening and statin use for some patients who meet age and cardiovascular disease risk criteria, BMI is not one of those criteria [31,32]. This study identified a cut point of 27.1 kg/m^2 for hyperlipidemia risk, indicating that the inclusion of BMI as a risk factor may be warranted. The USPSTF does not recommend screening for CAD but suggests that clinicians offer or refer adults with a BMI ≥30 kg/m^2 for behavioral weight loss therapy to prevent CAD development [33]. We are not aware of any USPSTF or professional society screening recommendations for osteoarthritis. Screening questionnaires for osteoarthritis exist [34] and could be provided to patients who exceed the BMI cut point of 28.7 kg/m^2. We also identified sex and race or ethnicity differences that may need to be considered when screening adults for obesity-related comorbidities.

Our previous EHR publication found that our patient population was demographically similar to the US adult population [35]; thus, our findings may be generalizable to US adults. However, further investigation of the BMI cut points identified in this study using multi-institutional EHR data sets would further elucidate whether our findings are generalizable. If the BMI cut points are similar within multi-institutional EHR data sets, screening recommendations for some comorbidities may need to be re-evaluated to help guide health care providers on when to screen patients for obesity-related comorbidities.

Limitations

First, although our methodology using the Youden index is established in the literature [6,19], there is no *gold standard* method for determining optimal cut points for continuous data, such as BMI. Some investigators have used disease prevalence rather than incidence to establish cut points. Our analysis comparing cut point calculations using incidence versus prevalence identified no clinically significant differences. We believe that cut points determined with incidence have more clinical utility because incidence evaluates the development of disease, whereas prevalence describes a disease that has already been diagnosed. Second, most Youden indices, sensitivities, and specificities were low, which suggests that BMI is not a perfect screening tool for these diseases. However, it has significant clinical use because it is recorded for most patients in the EHR, whereas other markers, such as waist circumference and biomarkers, are not. In addition, the AUROCs were >0.6, indicating that our analyses were able to discriminate between those with and without the disease. Third, there may be selection bias, given that all patients were required to have data in our EHR spanning at least 2 years. For example, our EHR had a lower percentage of Medicaid patients than the national estimates. Fourth, our study was observational, so no inferences can be made about causation. Finally, there may be inaccuracies in our data set because of errors in data entry by health care providers. We removed biologically implausible values using our BMI algorithm, but coding inaccuracies in the ICD-9 and ICD-10 entries may still exist.

Conclusions

The BMI cut points that accurately predict the risks of developing six obesity-related comorbidities (CAD, hyperlipidemia, hypertension, OSA, osteoarthritis, and T2DM) occurred when patients were overweight or barely met the criteria for class 1 obesity. Weight loss counseling for these patients is important because they are at an increased risk of morbidity and mortality related to obesity. Further studies using longitudinal, national data are needed to determine whether screening guidelines for CAD, hyperlipidemia, OSA, and osteoarthritis should be reconsidered.

Acknowledgments

Effort on this study and manuscript was made possible by an American College of Surgeons George HA Clowes Career Development Award, and a Veterans Affair (VA) Career Development Award to LF (CDA 015-060). The views represented in this study represent those of the authors and not those of the Department of Veteran Affairs or the US government. The project described was also supported by the Clinical and Translational Science Award program through the National Institutes of Health (NIH) National Center for Advancing Translational Sciences (grant UL1TR002373). Further funding was obtained through the NIH T32 Surgical Oncology Research Training Program (grant T32 CA090217-17). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. GC was partially funded through the Patient‐Centered Outcomes Research Institute Awards (ME-2018C2-13180). The views in this publication are solely the responsibility of the authors and do not necessarily represent the views of the Patient‐Centered Outcomes Research Institute, its Board of Governors, or Methodology Committee. The authors would like to acknowledge the internal fund from the Department of Family Medicine and Community Health for supporting GC’s effort on the collaboration between the Department of Family Medicine and Community Health and the Department of Biostatistics and Medical Informatics. This study was presented as an oral presentation at the American College of Surgeons Clinical Congress 2019 on October 29, 2019, in San Francisco, California, and as an oral presentation at the 2019 Wisconsin Surgical Society meeting on November 8, 2019, in Kohler, Wisconsin.

Authors' Contributions

LF, NL, JB, MV, LH, and GC contributed to the study design. NL, JB, MV, and GC contributed to data collection and analysis. NL, JB, and LF contributed to manuscript composition. All coauthors participated in data interpretation and manuscript revision. All coauthors approved the version to be published and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work were appropriately investigated and resolved.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Strengthening the Reporting of Observational Studies in Epidemiology statement checklist.
[DOCX File , 32 KB - jmir_v23i8e24017_app1.docx]

Multimedia Appendix 2
International Classification of Disease-9 and International Classification of Disease-10 codes used to identify comorbidities.
[DOCX File , 14 KB - jmir_v23i8e24017_app2.docx]

Multimedia Appendix 3
The 1-year incidence rates of comorbidities.
[DOCX File , 14 KB - jmir_v23i8e24017_app3.docx]

Multimedia Appendix 4
Quantile regression analysis of the associations between the incidence of comorbidities and median BMIs.
[DOCX File , 14 KB - jmir_v23i8e24017_app4.docx]

Multimedia Appendix 5
Comparison of baseline characteristics between patients who developed coronary artery disease and those who did not.
[DOCX File , 17 KB - jmir_v23i8e24017_app5.docx]

Multimedia Appendix 6
Comparison of baseline characteristics between patients who developed hyperlipidemia and those who did not.
[DOCX File , 16 KB - jmir_v23i8e24017_app6.docx]

Multimedia Appendix 7
Comparison of baseline characteristics between patients who developed hypertension and those who did not.
[DOCX File , 16 KB - jmir_v23i8e24017_app7.docx]

Multimedia Appendix 8
Comparison of baseline characteristics between patients who developed obstructive sleep apnea and those who did not.
[DOCX File , 16 KB - jmir_v23i8e24017_app8.docx]

Multimedia Appendix 9
Comparison of baseline characteristics between patients who developed osteoarthritis and those who did not.
[DOCX File , 16 KB - jmir_v23i8e24017_app9.docx]

Multimedia Appendix 10
Comparison of baseline characteristics between patients who developed type 2 diabetes mellitus and those who did not.
[DOCX File , 16 KB - jmir_v23i8e24017_app10.docx]

Multimedia Appendix 11
Area under the receiving operating curve, Youden index, and sensitivity or specificity of sex and race or ethnicity-specific cut points.
[DOCX File , 21 KB - jmir_v23i8e24017_app11.docx]

Multimedia Appendix 12
Incidence versus prevalence cut points.
[DOCX File , 13 KB - jmir_v23i8e24017_app12.docx]

References

1. Ward ZJ, Bleich SN, Cradock AL, Barrett JL, Giles CM, Flax C, et al. Projected U.S. state-level prevalence of adult obesity and severe obesity. N Engl J Med 2019 Dec 19;381(25):2440-2450. [doi: 10.1056/nejmsa1909301]
2. Mascie-Taylor CGN, Goto R. Human variation and body mass index: a review of the universality of BMI cut-offs, gender and urban-rural differences, and secular changes. J Physiol Anthropol 2007 Mar;26(2):109-112 [FREE Full text] [doi: 10.2114/jpa2.26.109] [Medline: 17435352]
3. Obesity : Preventing and Managing the Global Epidemic : Report of a WHO Consultation on Obesity. Geneva: World Health Organization; 1998:1-178.
4. WHO Expert Consultation. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. Lancet 2004 Jan;363(9403):157-163. [doi: 10.1016/s0140-6736(03)15268-3]
5. Arroyo-Johnson C, Mincey K. Obesity epidemiology worldwide. Gastroenterol Clin North Am 2016 Dec;45(4):571-579 [FREE Full text] [doi: 10.1016/j.gtc.2016.07.012] [Medline: 27837773]
6. Chu F, Hsu C, Jeng C. Lowered cutoff points of obesity indicators are better predictors of hypertension and diabetes mellitus in premenopausal Taiwanese women. Obes Res Clin Pract 2015 Jul;9(4):328-335. [doi: 10.1016/j.orcp.2014.08.001] [Medline: 25154649]
7. Jih J, Mukherjea A, Vittinghoff E, Nguyen TT, Tsoh JY, Fukuoka Y, et al. Using appropriate body mass index cut points for overweight and obesity among Asian Americans. Prev Med 2014 Aug;65:1-6 [FREE Full text] [doi: 10.1016/j.ypmed.2014.04.010] [Medline: 24736092]
8. Pantalone KM, Hobbs TM, Chagin KM, Kong SX, Wells BJ, Kattan MW, et al. Prevalence and recognition of obesity and its associated comorbidities: cross-sectional analysis of electronic health record data from a large US integrated health system. BMJ Open 2017 Nov 16;7(11):e017583 [FREE Full text] [doi: 10.1136/bmjopen-2017-017583] [Medline: 29150468]
9. Guh DP, Zhang W, Bansback N, Amarsi Z, Birmingham CL, Anis AH. The incidence of co-morbidities related to obesity and overweight: a systematic review and meta-analysis. BMC Public Health 2009 Mar 25;9(1):88 [FREE Full text] [doi: 10.1186/1471-2458-9-88] [Medline: 19320986]
10. Faeh D, Braun J, Tarnutzer S, Bopp M. Obesity but not overweight is associated with increased mortality risk. Eur J Epidemiol 2011 Aug 17;26(8):647-655. [doi: 10.1007/s10654-011-9593-2] [Medline: 21681546]
11. Klompas M, Cocoros NM, Menchaca JT, Erani D, Hafer E, Herrick B, et al. State and local chronic disease surveillance using electronic health record systems. Am J Public Health 2017 Sep;107(9):1406-1412. [doi: 10.2105/ajph.2017.303874]
12. Fildes A, Charlton J, Rudisill C, Littlejohns P, Prevost AT, Gulliford MC. Probability of an obese person attaining normal body weight: cohort study using electronic health records. Am J Public Health 2015 Sep;105(9):54-59. [doi: 10.2105/ajph.2015.302773]
13. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol 2008 Apr;61(4):344-349. [doi: 10.1016/j.jclinepi.2007.11.008] [Medline: 18313558]
14. Cheng FW, Gao X, Mitchell DC, Wood C, Still CD, Rolston D, et al. Body mass index and all-cause mortality among older adults. Obesity (Silver Spring) 2016 Oct 29;24(10):2232-2239 [FREE Full text] [doi: 10.1002/oby.21612] [Medline: 27570944]

15. Liu N, Birstler J, Venkatesh M, Hanrahan L, Chen G, Funk L. Weight loss for patients with obesity: an analysis of long-term electronic health record data. Med Care 2020 Mar;58(3):265-272 [FREE Full text] [doi: 10.1097/MLR.0000000000001277] [Medline: 31876663]
16. Nguyen NT, Magno CP, Lane KT, Hinojosa MW, Lane JS. Association of hypertension, diabetes, dyslipidemia, and metabolic syndrome with obesity: findings from the National Health and Nutrition Examination Survey, 1999 to 2004. J Am Coll Surg 2008 Dec;207(6):928-934. [doi: 10.1016/j.jamcollsurg.2008.08.022] [Medline: 19183541]
17. Zhao G, Ford ES, Dhingra S, Li C, Strine TW, Mokdad AH. Depression and anxiety among US adults: associations with body mass index. Int J Obes (Lond) 2009 Feb 6;33(2):257-266. [doi: 10.1038/ijo.2008.268] [Medline: 19125163]
18. Koenker R, Hallock KF. Quantile regression. J Econ Perspect 2001 Nov 01;15(4):143-156. [doi: 10.1257/jep.15.4.143]
19. Tseng C, Chong C, Chan T, Bai C, You S, Chiou H, et al. Optimal anthropometric factor cutoffs for hyperglycemia, hypertension and dyslipidemia for the Taiwanese population. Atherosclerosis 2010 Jun;210(2):585-589. [doi: 10.1016/j.atherosclerosis.2009.12.015] [Medline: 20053403]
20. Fujita M, Hata A. Sex and age differences in the effect of obesity on incidence of hypertension in the Japanese population: a large historical cohort study. J Am Soc Hypertens 2014 Jan;8(1):64-70. [doi: 10.1016/j.jash.2013.08.001] [Medline: 24113321]
21. Ong KL, Tso AW, Lam KS, Cheung BM. Gender difference in blood pressure control and cardiovascular risk factors in Americans with diagnosed hypertension. Hypertension 2008 Apr;51(4):1142-1148. [doi: 10.1161/hypertensionaha.107.105205]
22. Fontaine KR, Redden DT, Wang C, Westfall AO, Allison DB. Years of life lost due to obesity. J Am Med Assoc 2003 Jan 08;289(2):187-193. [doi: 10.1001/jama.289.2.187] [Medline: 12517229]
23. Wagner D, Heyward V. Measures of body composition in blacks and whites: a comparative review. Am J Clin Nutr 2000 Jun;71(6):1392-1402. [doi: 10.1093/ajcn/71.6.1392] [Medline: 10837277]
24. Cheong KC, Yusoff AF, Ghazali SM, Lim KH, Selvarajah S, Haniff J, et al. Optimal BMI cut-off values for predicting diabetes, hypertension and hypercholesterolaemia in a multi-ethnic population. Public Health Nutr 2012 Jun 01;16(3):453-459. [doi: 10.1017/s1368980012002911]
25. Chan P, Tai B, Loo G, Koo C, Ong T, Yeo T, et al. Optimal body mass index cut-offs for identification of patients with coronary artery disease at high risk of obstructive sleep apnoea. Heart Lung Circ 2016 Aug;25(8):847-854. [doi: 10.1016/j.hlc.2016.02.010] [Medline: 27067667]
26. Siu AL. Screening for abnormal blood glucose and type 2 diabetes mellitus: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med 2015 Oct 27;163(11):861. [doi: 10.7326/m15-2345]
27. Siu AL. Screening for high blood pressure in adults: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med 2015 Oct 13;163(10):778. [doi: 10.7326/m15-2223]
28. Aurora RN, Quan SF. Quality measure for screening for adult obstructive sleep apnea by primary care physicians. J Clin Sleep Med 2016 Aug 15;12(8):1185-1187 [FREE Full text] [doi: 10.5664/jcsm.6064] [Medline: 27397659]
29. Squire P. Obstructive sleep apnea. In: Goldschneider K, Davidson A, Wittkugel E, Skinner A, editors. Clinical Pediatric Anesthesia: A Case-Based Handbook. Oxford, United Kingdom: Oxford University Press; 2013.
30. Final report: Obstructive sleep apnea (Task 11-05). Federal Motor Carrier Safety Administration. 2005. URL: https://www.fmcsa.dot.gov/advisory-committees/mcsac/final-report-obstructive-sleep-apnea-task-11-05 [accessed 2019-08-01]
31. US Preventive Services Task Force. Screening adults for lipid disorders: recommendations and rationale. Am J Prev Med 2001 Apr;20(3 Suppl):73-76. [Medline: 11306235]
32. Sniderman AD. Did the ACC/AHA/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA cholesterol guidelines get apoB right? J Clin Lipidol 2019 May;13(3):360-366. [doi: 10.1016/j.jacl.2019.05.010] [Medline: 31229023]
33. US Preventive Services Task Force, Curry SJ, Krist AH, Owens DK, Barry MJ, Caughey AB, et al. Behavioral weight loss interventions to prevent obesity-related morbidity and mortality in adults: US Preventive Services Task Force recommendation statement. J Am Med Assoc 2018 Sep 18;320(11):1163-1171. [doi: 10.1001/jama.2018.13022] [Medline: 30326502]
34. Roux CH, Saraux A, Mazieres B, Pouchot J, Morvan J, Fautrel B, KHOALA Osteoarthritis Group. Screening for hip and knee osteoarthritis in the general population: predictive value of a questionnaire and prevalence estimates. Ann Rheum Dis 2008 Oct 01;67(10):1406-1411. [doi: 10.1136/ard.2007.075952] [Medline: 18077540]
35. Funk L, Shan Y, Voils C, Kloke J, Hanrahan L. Electronic health record data versus the National Health and Nutrition Examination Survey (NHANES): a comparison of overweight and obesity rates. Med Care 2017 Jun;55(6):598-605 [FREE Full text] [doi: 10.1097/MLR.0000000000000693] [Medline: 28079710]

Abbreviations

AUROC: area under the receiver operating curve
CAD: coronary artery disease
EHR: electronic health record
ICD: International Classification of Disease
NHANES: National Health and Nutrition Examination Survey

NIH: National Institutes of Health
OSA: obstructive sleep apnea
T2DM: type 2 diabetes mellitus
USPSTF: United States Preventative Services Task Force

Edited by R Kukafka, G Eysenbach; submitted 31.08.20; peer-reviewed by A Jo, J Lau; comments to author 25.09.20; revised version received 01.11.20; accepted 21.06.21; published 09.08.21.

Please cite as:
Liu N, Birstler J, Venkatesh M, Hanrahan L, Chen G, Funk L
Obesity and BMI Cut Points for Associated Comorbidities: Electronic Health Record Study
J Med Internet Res 2021;23(8):e24017
URL: https://www.jmir.org/2021/8/e24017
doi:10.2196/24017
PMID:34383661

©Natalie Liu, Jen Birstler, Manasa Venkatesh, Lawrence Hanrahan, Guanhua Chen, Luke Funk. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 09.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

A Fine-Tuned Bidirectional Encoder Representations From Transformers Model for Food Named-Entity Recognition: Algorithm Development and Validation

Riste Stojanov[1*], PhD; Gorjan Popovski[2,3*], MSc; Gjorgjina Cenikj[2,3*], BSc; Barbara Koroušić Seljak[2*], PhD; Tome Eftimov[2*], PhD

[1]Faculty of Computer Science and Engineering, Ss Cyril and Methodius, University- Skopje, Skopje, the Former Yugoslav Republic of Macedonia
[2]Computer Systems Department, Jožef Stefan Institute, Ljubljana, Slovenia
[3]Jožef Stefan International Postgraduate School, Ljubljana, Slovenia
[*]all authors contributed equally

Corresponding Author:
Tome Eftimov, PhD
Computer Systems Department
Jožef Stefan Institute
Jamova Cesta 39
Ljubljana, 1000
Slovenia
Phone: 386 1 477 3386
Email: tome.eftimov@ijs.si

Abstract

Background: Recently, food science has been garnering a lot of attention. There are many open research questions on food interactions, as one of the main environmental factors, with other health-related entities such as diseases, treatments, and drugs. In the last 2 decades, a large amount of work has been done in natural language processing and machine learning to enable biomedical information extraction. However, machine learning in food science domains remains inadequately resourced, which brings to attention the problem of developing methods for food information extraction. There are only few food semantic resources and few rule-based methods for food information extraction, which often depend on some external resources. However, an annotated corpus with food entities along with their normalization was published in 2019 by using several food semantic resources.

Objective: In this study, we investigated how the recently published bidirectional encoder representations from transformers (BERT) model, which provides state-of-the-art results in information extraction, can be fine-tuned for food information extraction.

Methods: We introduce FoodNER, which is a collection of corpus-based food named-entity recognition methods. It consists of 15 different models obtained by fine-tuning 3 pretrained BERT models on 5 groups of semantic resources: food versus nonfood entity, 2 subsets of Hansard food semantic tags, FoodOn semantic tags, and Systematized Nomenclature of Medicine Clinical Terms food semantic tags.

Results: All BERT models provided very promising results with 93.30% to 94.31% macro F1 scores in the task of distinguishing food versus nonfood entity, which represents the new state-of-the-art technology in food information extraction. Considering the tasks where semantic tags are predicted, all BERT models obtained very promising results once again, with their macro F1 scores ranging from 73.39% to 78.96%.

Conclusions: FoodNER can be used to extract and annotate food entities in 5 different tasks: food versus nonfood entities and distinguishing food entities on the level of food groups by using the closest Hansard semantic tags, the parent Hansard semantic tags, the FoodOn semantic tags, or the Systematized Nomenclature of Medicine Clinical Terms semantic tags.

(J Med Internet Res 2021;23(8):e28229) doi:10.2196/28229

KEYWORDS

food information extraction; named-entity recognition; fine-tuning BERT; semantic annotation; information extraction; BERT; bidirectional encoder representations from transformers; natural language processing; machine learning

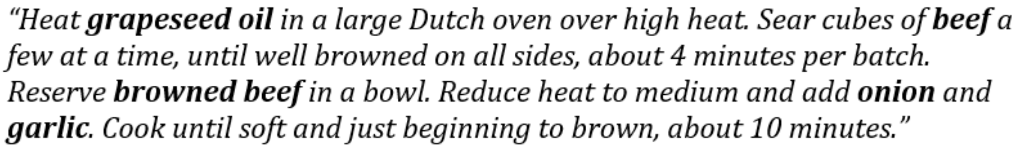
Introduction

Food is one of the most important environmental factors that affects human health [1]. However, even healthy and ecofriendly foods can cause health problems when consumed together with specific drugs or while having specific diseases. Comprehensive dietary assessments are required to understand how food influences our health, after considering various aspects. Automating the detection of food entities is important for several applications such as food-drug interactions and health issues related to food.

Computer science can greatly contribute to this research topic, especially in the areas of machine learning, natural language processing (NLP), and data analysis. Data collected in studies carry important information, which is not easily extracted when it has been gathered from different data sources. The main problem is that these data are presented in different formats: structured, semistructured, and unstructured. Additionally, the data consist of entities from different domains such as food and nutrition, medicine, pharmacy, ecology, and agriculture. The extraction of this information allows the creation of knowledge graphs [2], which represent a collection of interlinked descriptions of entities—objects, events, or concepts—by using semantic metadata and providing a framework for data integration, unification, analytics, and sharing.

To create a knowledge graph, first, we should have methods that can be used for information extraction, which is the task of automatically extracting structured information from unstructured textual data. In most cases, information extraction is performed by using named-entity recognition (NER) methods (ie, a subtask of information extraction), which deal with automatically detecting and identifying phrases (ie, one or more words [tokens]) from the text that represents the domain entities. Let us assume the following recipe example (Figure 1):

Figure 1. Recipe example.



"Heat ***grapeseed oil*** *in a large Dutch oven over high heat. Sear cubes of* ***beef*** *a few at a time, until well browned on all sides, about 4 minutes per batch. Reserve* ***browned beef*** *in a bowl. Reduce heat to medium and add* ***onion*** *and* ***garlic****. Cook until soft and just beginning to brown, about 10 minutes."*

The phrases in bold (Figure 1) are the named entities that should be recognized in the process of information extraction, and they should be linked to their corresponding domain entity tag. In the simplest case, they may be linked to the generic "Food" class, but extracting the more specific food class by a level of food group may be of higher value, because this class may potentially provide multiple nutrition facts that may allow new use cases such as ingredient substitution.

Several types of NER methods exist depending on their underlying methodology: (1) *dictionary-based* [3], which return only entities that are mentioned in the dictionary in which they are based; (2) *rule-based* [4,5], which use a dictionary in combination with rules that describe the characteristics of the entities in the domain of interest; (3) *corpus-based* [6,7], which learn a supervised machine learning model by using an annotated corpus; (4) *active learning–based* [8], which use semisupervised learning to train a model that does not require a large annotated corpus but instead interacts with the user to query for new annotations that are used for iteratively improving the model; and (5) *deep learning–based* [9], which use deep neural networks to train a model that requires a large amount of annotated data. Nowadays, fine-tuning the bidirectional encoder representations from transformers (BERT) [10] provides state-of-the-art results in NER tasks. However, the task of fine-tuning the BERT model for NER requires a domain-specific annotated corpus.

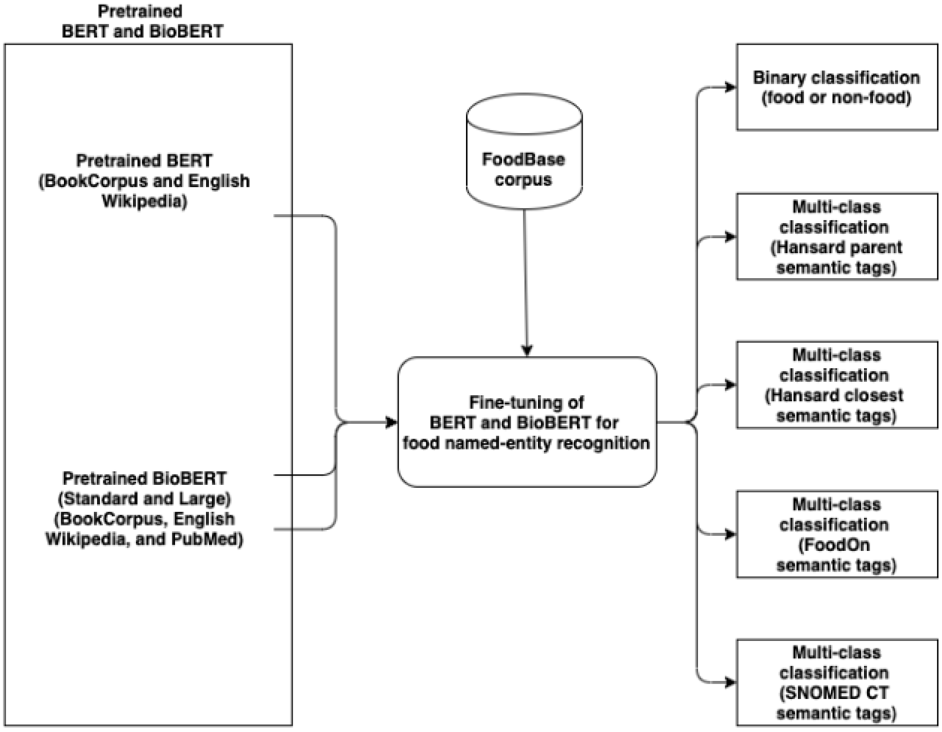
In the past 2 decades, a large amount of work has been done to address this problem in the biomedical domain [11-17]. All of this work is supported by the existence of diverse biomedical vocabularies and standards such as the Unified Medical Language System [18], together with the collection of a large amount of annotated biomedical data (eg, in the domain of drugs, diseases, and other treatments) from numerous biomedical NLP workshops [19-26]. The existence of such resources and information extraction methods allows the creation of knowledge graphs that can support the biomedical domain and clinical practices [27,28].

In contrast to the biomedical domain, the food domain is relatively inadequately resourced. There are few semantic models (ie, ontologies) [29], each of which has been developed for very specific applications. One such example is the Ontology for Nutritional Epidemiology, which was developed to describe dietary food assessment [30]. Until recently, there was no annotated food corpus, which meant that the available food NERs were rule-based. Hanisch et al [4] presented a rule-based NER known as drNER for information extraction from evidence-based dietary recommendations. Food entities are among the domain entities of interest that are extracted. However, drNER extracts several food entities as one. This was improved by developing the rule-based NER Food Information Extraction [31], where the rules incorporate computational linguistics information in combination with food semantic annotations from the Hansard corpus [32]. Another way to perform food information extraction is to use the NCBO (National Center for Biomedical Ontology) annotator [33], which is a web service that annotates text by using food ontology concepts that are part of the BioPortal software services [34]. It can be combined with the following ontologies: FoodOn [35], OntoFood, and SNOMED CT (Systematized Nomenclature of Medicine Clinical Terms) [36]. A comparison of 4 NER methods (Food Information Extraction, NCBO [SNOMED CT], NCBO [OntoFood], and NCBO [FoodOn]) is presented by Popovski et al [37], who showed that Food Information Extraction provides the best results in distinguishing food from nonfood entities. The main weakness of the abovementioned NERs is that they all depend on other external resources such as taxonomies, ontologies, or previously developed annotators, which further can be a problem if some of the resources become inaccessible. This also opens new directions for future research regarding the development of more robust food NERs.

At the end of 2019, an annotated food corpus known as FoodBase [38] was published. The ground truth corpus consists of 1000 recipes, where for each recipe, the food entities mentioned in it are first extracted and then annotated using the hierarchical Hansard food semantic tags (eg, AG.01 [food], AG.01.h.02 [vegetables], AG.01.h.02.i [herb], AG.01.n.15 [pastry], AE.10 [fish]). The corpus is organized according to the BioC format, which is a minimalist approach for interoperability for biomedical text processing [39]. The availability of the FoodBase corpus allowed the development of the first food corpus-based NER known as bidirectional long short-term memory for food named-entity recognition (BuTTER) [40], where bidirectional long short-term memory (BiLSTM) in conjunction with conditional random fields (CRFs) and different representation learning methods have been explored to develop NER that distinguishes between food versus nonfood entities. In addition to this, the FoodOntoMap resource was published [41], where for the same entities found in FoodBase, the semantic tags from FoodOn, OntoFood, and SNOMED CT were assigned. With this, the food entities were normalized to different food semantic resources, which additionally links the food semantic resources.

Enabled by the availability of several food resources that were published toward the end of 2019, we introduce a fine-tuned BERT model that can be used for food information extraction, called as FoodNER. BERT is known to achieve state-of-the-art results in NER tasks [42-44], and hence, we utilize it to develop a more robust model for food information extraction. The flowchart of FoodNER is presented in Figure 2. It is developed using a predefined BERT model, which can be the original BERT or some variation of BioBERT. Using them, fine-tuning is performed on the FoodBase corpus to address several different tasks: food or nonfood entity and 4 types of distinguishing food entities, depending on the semantic resource from where the semantic tags are taken (ie, Hansard semantic taxonomy [done twice on different hierarchical levels from the taxonomy], FoodOn, and SNOMED CT).

Figure 2. Food named-entity recognition flowchart. BERT: bidirectional encoder representations from transformers; NER: named-entity recognition; SNOMED CT: Systematized Nomenclature of Medicine Clinical Terms.


Pretrained
BERT and BioBERT
Pretrained BERT
(BookCorpus and English
Wikipedia)
Pretrained BioBERT
(Standard and Large)
(BookCorpus, English
Wikipedia, and PubMed)
FoodBase
corpus
Fine-tuning of
BERT and BioBERT for
food named-entity recognition
Binary classification
(food or non-food)
Multi-class
classification
(Hansard parent
semantic tags)
Multi-class
classification
(Hansard closest
semantic tags)
Multi-class
classification
(FoodOn
semantic tags)
Multi-class
classification
(SNOMED CT
semantic tags)

The main contributions of this study are as follows:

1. We fine-tuned different BERT models on different semantic resources from which the food semantic tags are taken. All BERT models have very promising results, obtaining around 73.39%-78.96% macro F1 score. All in all, it represents the new state-of-the-art in food information extraction.
2. In comparison with the already existing food rule–based (Food Information Extraction) and corpus-based (BuTTER) NER methods regarding the task of distinguish between food or nonfood entity, FoodNER provides similar results. However, it is more robust than the rule-based approaches since it does not require the continuous availability of additional external resources, which can be a problem regarding sustainability. Additionally, comparing it to the corpus-based method BuTTER, it is the first model that can predict food groups instead of just distinguishing between food versus nonfood entities.
3. The source code used for fine-tuning the different FoodNER models is publicly available. All models are also included in FoodViz [45], which is a new tool for the visualization

of food annotations in text. The users can additionally select which model they want to use and annotate their data.

In this study, we used the FoodBase ground truth corpus for building and evaluating FoodNER models for distinguishing food versus nonfood entity as well as for distinguishing food entities concerning the Hansard semantic tags. The BuTTER approach is used as baseline for comparing the performance of the FoodNER models. The FoodOntoMap extension of the FoodBase ground truth corpus is also used for training and evaluating the FoodNER models concerning the SNOMED CT and FoodOn semantic tags.

Methods

FoodBase Data Corpus

The FoodBase data corpus is a recently published corpus with food annotations [38]. It consists of 2 versions: curated and uncurated. The curated version consists of 1000 recipes that are annotated using a rule-based NER and then manually checked by subject matter experts who removed the false positives and added the false negatives to create a ground truth standard. It consists of 200 recipes for each of the following recipe categories: appetizers and snacks, breakfast and lunch, dessert, dinner, and drinks. The uncurated version consists of approximately 22,000 recipes, which are only annotated with the rule-based NER, without being checked by subject matter experts. The semantic tags used for annotations are taken from the Hansard corpus [32,45]. To the best of our knowledge, this is the first corpus with such annotated food entities.

Food Semantic Resources

Hansard Corpus

The Hansard corpus [32] is part of the SAMUELS (Semantic Annotation and Mark-Up for Enhancing Lexical Searches) project, where semantic tags are organized in a hierarchy with 37 higher-level semantic groups. One of these groups is the *Food and Drink*, which is then split into 3 subcategories, that is, *food, production of food, farming,* and *acquisition of animals for food, hunting*. These have 125, 36, and 13 top-level semantic tags, respectively.

FoodOn Ontology

FoodOn is a farm-to-fork ontology about food, which supports food traceability [35]. It consists of information about food products, their sources, and information about preservation processes, packaging, etc. It is built to represent food-related entities and to provide vocabulary for nutrition, diet, and plant and animal agricultural rearing research. FoodOn interoperates with the Open Biological and Biomedical Ontology Library and imports material from several ontologies covering anatomy, taxonomy, geography, and cultural heritage. The ontology aims to cover gaps in the representation of food-related products and processes and is being applied to research and clinical data sets in the academia and government.

SNOMED CT Ontology

SNOMED CT is the most comprehensive multilingual clinical health care terminology [36]. It is a machine-readable collection of medical terms, where synonyms and clinical definitions are available for each of the codes. It consists of information about drugs, disorders, symptoms, diagnoses, procedures, body structures, food, and other concepts that are related to health care.

FoodOntoMap

FoodOntoMap is a recently published resource that is developed by using the FoodBase corpus [38]. It provides data normalization of the food entities according to different semantic resources. Specifically, for each extracted entity presented in the FoodBase corpus, the semantic tags from Hansard, FoodOn, OntoFood, and SNOMED CT are available. It is important to note that the semantic tags from resources other than Hansard are not available for some of the extracted food entities since they do not exist in the respective food ontologies themselves. The food entity coverage per semantic resource is presented by Popovski et al [37].

BERT

BERT is a word representation model that achieves state-of-the-art results in many NLP tasks [10]. The main idea of BERT is the bidirectional training of the transformer, which is different from previously published models that were trained using just a text sequence either from left to right or from right to left. Many models predict the next word in a sequence, while BERT uses a masked model, which predicts words masked in random order. It is used for bidirectional representation learning. BERT follows the idea and value of transfer learning [46,47], starting with pretraining a representation language model and then performing fine-tuning of the model for a new learning task (eg, NER, Question Answering). The same architectures are used in the pretraining and the fine-tuning step. The only difference is in the output layers. The parameters from the pretrained model are used as initial parameters, which are further fine-tuned concerning the learning task that is being solved in the fine-tuning.

Pretraining of BERT

In this phase, we did not pretrain a BERT model on our corpus. Instead, we used 3 previously pretrained and publicly available BERT models to fine-tune them for the food NER task. Specifically, the 3 BERT models that were used were the original pretrained BERT model [10], the pretrained BioBERT standard model [15], and the BioBERT large model [15]. The original BERT model was trained on the BookCorpus with around 800 million words [48] and the English Wikipedia with around 2500 million words, from which only the texts were used, ignoring the headers, tables, and lists.

The BioBERT was trained to improve the model for tasks in the biomedical domain since the domain consists of a large number of domain-specific proper nouns and terms, which do not appear in normal texts. Different combinations of corpora were experimentally used for pretraining BioBERT. The combinations involved the following corpora: the BookCorpus and the English Wikipedia (same as the BERT model), PubMed abstracts with around 4500 million words, and PubMed Central full-text articles with around 13,500 million words. Finally, the model pretrained on the combination using the BookCorpus, the English Wikipedia, and PubMed abstracts using the BERT-base cased code provided by Google is known as the BioBERT language representation model (ie, BioBERT Standard). The same combination trained using the BERT-large cased code provided by Google is known as BioBERT large.

Fine-Tuning BERT

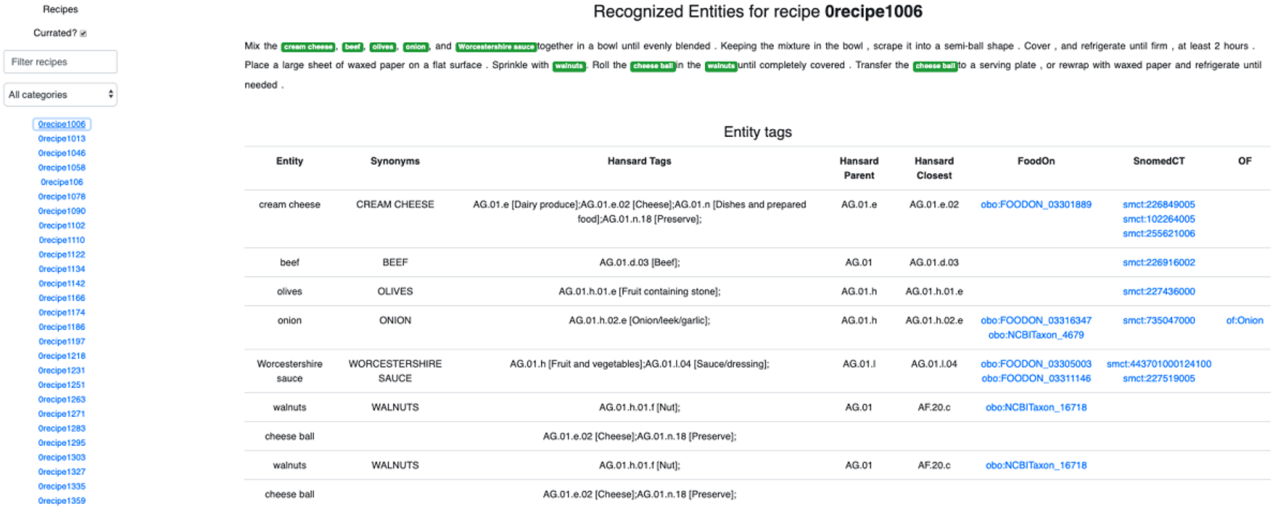
To perform food NER, we fine-tuned the original BERT and the 2 versions of the BioBERT model. In all the cases, for each class, we used the IOB (inside, outside, and beginning) tagging [49] prediction, which is a common tagging format in computational linguistics. In this process, we used the FoodBase corpus as the ground truth. However, this corpus may contain multiple Hansard tags for each food phrase, and we used a few methods for selecting the most representative tag for each phrase.

The fine-tuning was performed for the following tasks:

1. Food classification: This was performed for distinguishing food versus nonfood entity. In this task, we labeled all food phrases annotated in FoodBase with the tag FOOD and used this data set for training and validation.
2. Hansard parent: This was performed for distinguishing 48 classes from the Hansard corpus. In this task, we selected parent semantic tags from the Hansard hierarchy that correspond to the food phrases in FoodBase. In cases with multiple different parent tags present for the food phrase, we selected the first occurring parent.
3. Hansard closest: This was performed for distinguishing 92 classes from the Hansard hierarchy. In this task, for each food phrase in FoodBase, we chose the closest Hansard tag to the food phrase being annotated. The closest tag was selected using the minimum cosine distance between the BERT embedding of the food phrase and the BERT embeddings of the Hansard tag labels.
4. FoodOn: This was performed for distinguishing 205 classes, where the classes are semantic tags from the FoodOn ontology. For each food phrase in FoodBase, we selected the corresponding FoodOn class based on the FoodOntoMap mappings [40].
5. SNOMED CT: This was performed for distinguishing 207 classes, where the classes are semantic tags from the SNOMED CT ontology. In this task, we also used FoodOntoMap [40] to obtain the SNOMED CT class for the food phrase.

In cases of food versus nonfood entity task and the task of distinguishing food entities with regard to the Hansard semantic tags, we have a ground truth corpus—the curated part of FoodBase. However, in case of FoodOn and SNOMED CT, we fine-tuned BERT and BioBERT only for entities that had semantic tags provided by the FoodOntoMap resource (ie, not all food entities are presented in these 2 resources as was previously explained). All semantic tags (ie, Hansard parent, Hansard closest, FoodOn, and SNOMED CT) for each food entity available in the FoodBase corpus are presented by the FoodViz tool (see Figure 3). Finally, we ended up with 15 different fine-tuned models, 3 per task depending on the pretrained model that was used (BERT, BioBERT Standard, or BioBERT large).

Figure 3. An example of food entities available from one recipe that are present in the training data set. The entities are annotated using Hansard parent, Hansard closest, FoodOn, Systematized Nomenclature of Medicine Clinical Terms, and OntoFood (not studied in this paper) semantic tags.



Recognized Entities for recipe **0recipe1006**

Mix the cream cheese, beef, olives, onion, and Worcestershire sauce together in a bowl until evenly blended . Keeping the mixture in the bowl , scrape it into a semi-ball shape . Cover , and refrigerate until firm , at least 2 hours . Place a large sheet of waxed paper on a flat surface . Sprinkle with walnuts. Roll the cheese ball in the walnuts until completely covered . Transfer the cheese ball to a serving plate , or rewrap with waxed paper and refrigerate until needed .

Entity tags

Entity	Synonyms	Hansard Tags	Hansard Parent	Hansard Closest	FoodOn	SnomedCT	OF
cream cheese	CREAM CHEESE	AG.01.e [Dairy produce];AG.01.e.02 [Cheese];AG.01.n [Dishes and prepared food];AG.01.n.18 [Preserve];	AG.01.e	AG.01.e.02	obo:FOODON_03301889	smct:226849005 smct:102264005 smct:255621006	
beef	BEEF	AG.01.d.03 [Beef];	AG.01	AG.01.d.03		smct:226916002	
olives	OLIVES	AG.01.h.01.e [Fruit containing stone];	AG.01.h	AG.01.h.01.e		smct:227436000	
onion	ONION	AG.01.h.02.e [Onion/leek/garlic];	AG.01.h	AG.01.h.02.e	obo:FOODON_03316347 obo:NCBITaxon_4679	smct:735047000	of:Onion
Worcestershire sauce	WORCESTERSHIRE SAUCE	AG.01.h [Fruit and vegetables];AG.01.l.04 [Sauce/dressing];	AG.01.l	AG.01.l.04	obo:FOODON_03305003 obo:FOODON_03311146	smct:443701000124100 smct:227519005	
walnuts	WALNUTS	AG.01.h.01.f [Nut];	AG.01	AF.20.c	obo:NCBITaxon_16718		
cheese ball		AG.01.e.02 [Cheese];AG.01.n.18 [Preserve];					
walnuts	WALNUTS	AG.01.h.01.f [Nut];	AG.01	AF.20.c	obo:NCBITaxon_16718		
cheese ball		AG.01.e.02 [Cheese];AG.01.n.18 [Preserve];					

A Baseline for Comparison: BuTTER

To compare the results, the Bidirectional Long Short-Term Memory (LSTM) model for sequence tagging with a CRF layer (BiLSTM-CRF) [50] was used as a baseline, which has already been shown to achieve state-of-the-art results in several NLP tasks such as part-of-speech tagging, chunking, and NER tasks. Additionally, the BiLSTM-CRF model has been used to train food NER (food versus nonfood entity) utilizing the food annotations available in the FoodBase corpus [38], resulting in BuTTER models. BuTTER consists of 2 different BiLSTM-CRF architectures, each one evaluated with 3-word embedding methods (ie, GloVe [51], Word2Vec [52], and FastText [53]) and once using the word tokens for representing the textual data used by the input layer. The difference between the 2 BuTTER architectures is that the first one is a BiLSTM-CRF model without character embeddings, while the second one has an additional stacked input and embedding layer to generate character embeddings (Char-BiLSTM-CRF). When representing the textual data using the predefined vocabularies of Word2Vec, GloVe, and FastText, some of the words are absent; therefore, out-of-vocabulary word preprocessing techniques can be applied to handle them. In the case when word tokens are used, the impact of lemmatization on the model performance was investigated. All in all, results from 16 different BuTTER models were obtained, that is, 2 architectures × 4 textual representations (ie, 3-word embeddings + word tokens × 2 scenarios, that is, preprocessing applied or not). More details about them can be found in the study of Comeau et al [39].

Results

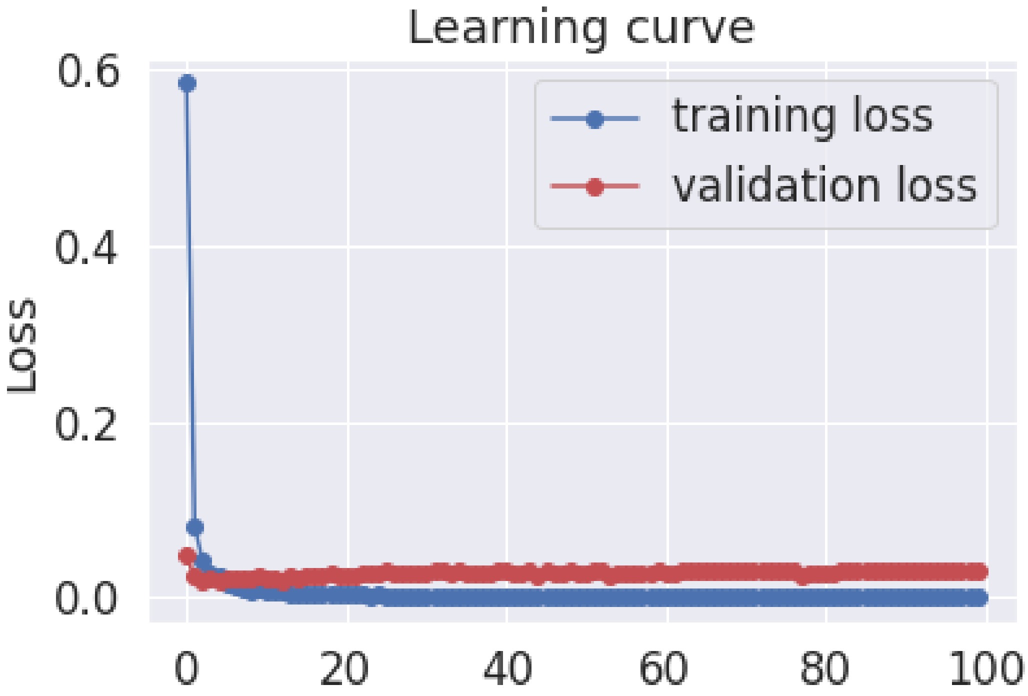
Experiments

In this section, the experimental setups for fine-tuning the BERT and BioBERT models in each classification task are explained, followed by the experimental results obtained by the evaluation. We performed 2 experiments: (1) comparison of the BERT models with the corpus-based BuTTER models presented in a previous study [40] on the food versus nonfood entity task, and (2) presenting results for BERT models that can distinguish between different food semantic tags.

Experimental Design

The experiments were performed using the Colab platform [54]. To fine-tune the pretrained BERT and BioBERT models, HuggingFace's *transformers* [55] library was used with its *BertForTokenClassification* class for token level prediction. This class wraps the *BertModel* class and adds a token-level classifier on top of it, which is a linear layer that takes the last hidden layer of the wrapped model as input. During the training of the fine-tuning, the *AdamW* optimizer was used with a *weight_decay_rate* of 0.01. The model was trained until its validation loss did not improve in 5 consecutive epochs, with a maximum of 100 epochs and with a scheduler to linearly reduce the learning rate throughout the epochs. Figure 4 presents the train and validation loss per fine-tuning epoch for the BioBERT large model on the Hansard parent data set. The same pattern holds for the other models, and therefore, we present the learning curve only for this particular model [56].

Figure 4. Training and validation loss per fine-tuning epoch for the bio bidirectional encoder representations from transformers large model on the Hansard parent data set.



For the BiLSTM-CRF model architecture or the BuTTER models, we used the default parameters presented in the study of Comeau et al [39], which are also presented here:

1. The maximum sequence length (ie, sentence length) is 50 since the longest sentence in the data set consists of 45 tokens.
2. The batch size is 256.
3. Architecture: input layer with 50 units, embedding layer with 300 units, BiLSTM layer with 50 units (total of 100 parameters), dense (TimeDistributed) layer with 50 units, CRF output layer where the final output dimension is the number of classes + 1 (ie, one for padding).

The aforementioned architecture refers to the complete architecture of the BuTTER BiLSTM-CRF model, that is, the model without character embeddings. The BuTTER Char-BiLSTM-CRF model contains an additional stack of input and embedding layers for generating the character embeddings and a concatenation layer for concatenating the word embeddings with the character embeddings. The additional input layer contains 18 units, while the additional embedding layer contains 20 units. Each of the BuTTER models was trained until the improvement in validation loss of 5 consecutive epochs did not surpass $5*10^{-3}$, to a maximum of 1000 epochs, whichever comes first.

The data sets used for training and testing are from the curated version of FoodBase [38] transformed in IOB tagging [49] format [57]. The train portion contains 81,347 tokens, while we report the results with the remaining 25,828 tokens, that is, approximately 75% of the data is used for training and the rest is for testing of the model. The curated version of FoodBase contains 1000 recipes, with 5 categories that contain 200 recipes each. We use 150 recipes with alphabetically smaller identifiers of each category for training and the rest of the recipes from the category for testing. The statistics about the number of tokens and their classes among the different data sets are shown in Table 1. The "Number of different inside, outside, and beginning annotations" row in this table describes the classes that our model tries to predict. Since we are predicting food phrases, for each different food phrase class, we may have annotations that start with *B-* for the first token in the phrase, and *I-* for all the rest of the tokens. Therefore, the number of different IOB annotations is approximately twice as large comparing it to the number of phrase classes. Additionally, the data sets are not balanced since the majority of the tokens are not part of the food phrase, that is, they are outside tokens. The *Hansard parent* data set is smaller than the others since there were 4 recipes with problematic parents and we omitted them in the evaluation.

Table 1. Data set statistics.

Annotations	Food classification	Hansard parent	Hansard closest	FoodOn	SNOMED CT[a]
Annotated tokens (beginning and inside)	17,937	11,759	17,864	8730	8151
Outside tokens	95,416	95,416	88,956	98,445	99,024
Number of different inside, outside, and beginning annotations	3	63	163	342	318
Number of food phrase classes	1	34	91	197	196
Total number of tokens	107,175	107,176	106,820	107,175	107,175

[a] SNOMED CT: Systematized Nomenclature of Medicine Clinical Terms.

The evaluation of the proposed models was done using stratified five-fold cross-validation. Stratified sampling was used to generate the folds since the FoodBase corpus consists of 5 different categories of recipes. For each recipe category, 10% of the training set of each fold was taken sequentially out and used for validation.

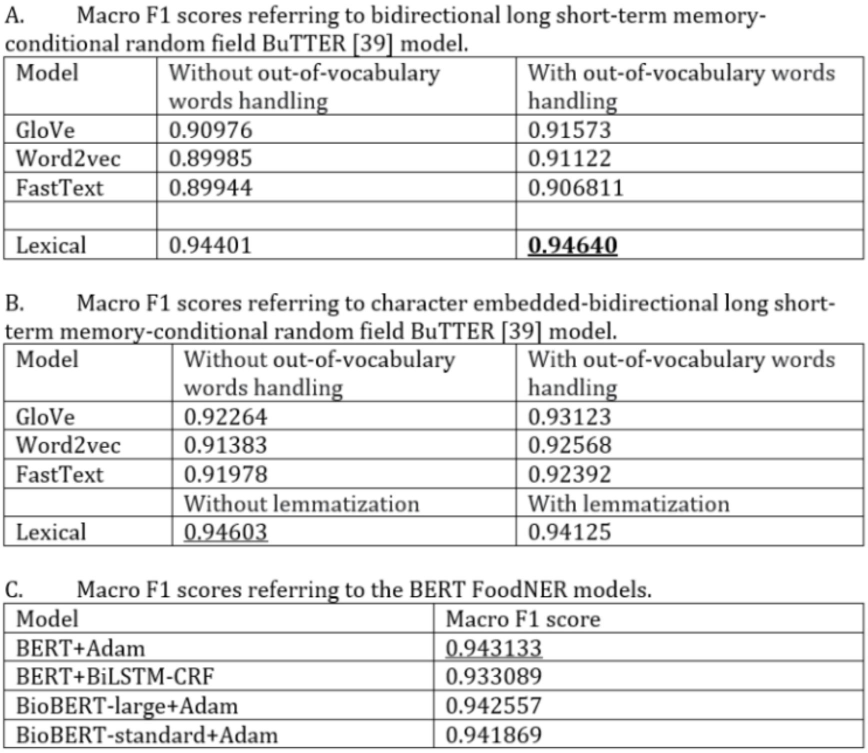
Experimental Results

Next, the results for both experiments are presented, starting with the comparison of the BERT models with the BuTTER models on the food versus nonfood task, followed by presenting the BERT models trained for distinguishing between different food semantic tags. We present the results for the macro F1 score. The macro averaging scheme computes each metric for each class independently and then calculates the mean. The rationale behind using macro averaging is that it conveys more meaningful information when considering especially a task that consists of more than two semantic tags that should be predicted with heavily unbalanced data. Conversely, simple micro averaging provides insufficient information in tasks where more than two semantic tags (ie, classes) are used, as it conflates the true positives, false positives, true negatives, and false negatives into one confusion matrix and then computes the evaluation metrics. Similarly, weighted averaging is biased in favor of the class most represented in the data, as the weight while computing the average depends on the relative frequency of the class label in the data set.

Comparison With the BuTTER Approach

Figure 5 [39] presents the results obtained from evaluating the fine-tuned BERT (ie, FoodNER) by using the original pretrained BERT model and 2 BioBERT models in the food versus nonfood task described in Methods and comparing them with the BuTTER results obtained for the same task. From the table, it follows that the best FoodNER model is obtained by fine-tuning the original pretrained BERT, resulting in a macro F1 score of 94.31%. Additionally, comparing it with the other FoodNER models obtained by fine-tuning BioBERT large and BioBERT standard, the absolute empirical differences are very small, amounting to only 0.05% and 0.12%, respectively. Comparing the FoodNER models with both BuTTER architectures (BiLSTM-CRF and Char-BiLSTM-CRF) when word embeddings are used to represent the textual data for the input layer (ie, GloVe, Word2Vec, and FastText), it follows that all FoodNER models have better macro F1 scores by using the stratified 5-fold cross-validation. However, we should point that the differences here are in the range from 1.74% to 4.36%. Comparing FoodNER models with both BuTTER architectures when word tokens are used for the input layer, the BiLSTM-CRF with lemmatization of the word tokens outperforms the FoodNER models by 0.32% and the Char-BiLSTM-CRF without lemmatization of the word tokens by 0.28%. We can conclude here that these differences are not crucial from a practical point of view; therefore, we can assume that all models perform similarly. Further, we also fine-tuned BERT by using the BiLSTM-CRF architecture for food classification, which results in a similar performance of a macro F1 score of 93.30%. To explore the robustness of the models, Figure 6 presents boxplots of the macro F1 score distributions obtained by evaluating each fold for each model separately. From the figure, it follows that all models perform well since all of them provide a macro F1 score greater than 87.00% for each fold. The most robust models are FoodNER BioBERT standard model and the BuTTER BiLSTM-CRF model with Word2Vec when out-of-vocabulary preprocessing is applied. However, comparing the results between both models, the FoodNER BioBERT standard provides a better macro F1 score. The other models also provide robust results, where the macro F1 scores obtained from different folds do no vary with large deviations. It is interesting to note that the best macro F1 score is obtained when BERT is fine-tuned with BiLST-CRF for one of the five folds; however, using the values from the other folds, the macro F1 score of this model can vary between different folds. Thus, we can conclude that the FoodNER models, which are fine-tuned BERT, BioBERT standard, and BioBERT large models, provide very robust results. These results also show that by using BERT, state-of-the-art results for food classification can be achieved.

Figure 5. Macro F1 scores for all considered models for the food versus nonfood entity task. Each macro F1 score is obtained by using stratified k-fold cross-validation (k=5). Underlined values are best per subtable, while the bold value is the best from the whole table. BERT: bidirectional encoder representations from transformers; BiLSTM-CRF: bidirectional long short-term memory conditional random field; BuTTER: bidirectional long short-term memory for food named-entity recognition; NER: named-entity recognition.



A. Macro F1 scores referring to bidirectional long short-term memory-conditional random field BuTTER [39] model.

Model	Without out-of-vocabulary words handling	With out-of-vocabulary words handling
GloVe	0.90976	0.91573
Word2vec	0.89985	0.91122
FastText	0.89944	0.906811
Lexical	0.94401	**0.94640**

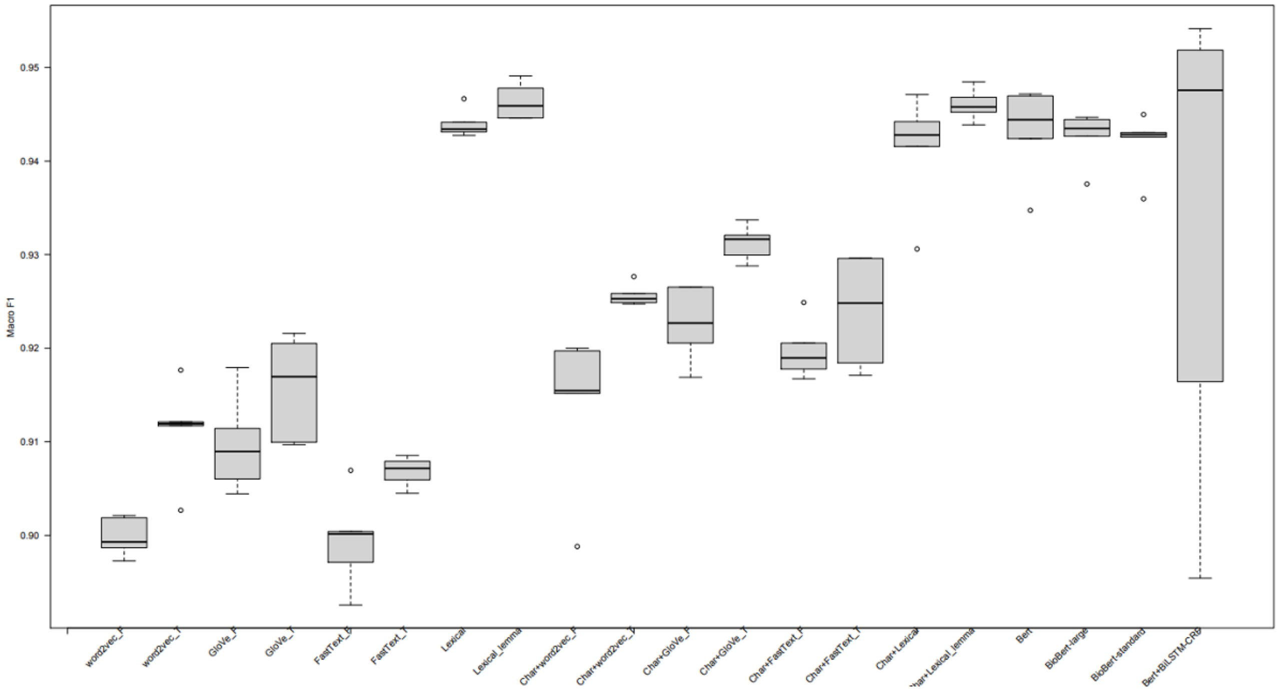
B. Macro F1 scores referring to character embedded-bidirectional long short-term memory-conditional random field BuTTER [39] model.

Model	Without out-of-vocabulary words handling	With out-of-vocabulary words handling
GloVe	0.92264	0.93123
Word2vec	0.91383	0.92568
FastText	0.91978	0.92392
	Without lemmatization	With lemmatization
Lexical	0.94603	0.94125

C. Macro F1 scores referring to the BERT FoodNER models.

Model	Macro F1 score
BERT+Adam	0.943133
BERT+BiLSTM-CRF	0.933089
BioBERT-large+Adam	0.942557
BioBERT-standard+Adam	0.941869

Figure 6. Boxplots of macro F1 scores obtained by using stratified five-fold cross-validation for all considered models for the binary food classification task. BERT: bidirectional encoder representations from transformers; BiLSTM-CRF: bidirectional long short-term memory conditional random field.



BERT Models for Recognizing Between Different Food Semantic Tags

In this experiment, we present the results of fine-tuning the BERT, BioBERT large, and BioBERT standard models in the tasks of distinguishing food entities concerning different semantic models (ie, FoodOn, Hansard closest, Hansard parent, and SNOMED CT). We have decided to focus only on the BERT models since it provides state-of-the-art results in already all NLP NER tasks. Additionally, in Table 1, the number of annotated tokens and the number of classes for each task are presented.

Table 2 provides the macro F1 scores for the 3 FoodNER models (BERT, BioBERT large, and BioBERT standard) for distinguishing food entities concerning different semantic models (ie, FoodOn, Hansard closest, Hansard parent, and SNOMED CT). The column *"epochs"* provides information for the number of epochs needed to fine-tune the model. From the table, it is evident that all models achieved a macro F1 score between 73.39% and 78.96%. The best models for each semantic tag set achieved the following macro F1 scores: (1) FoodOn, 78.13%; (2) Hansard closest, 78.96%; (3) Hansard parent, 76.26%; and (4) SNOMED CT, 76.01%.

Keeping in mind the number of classes we are predicting for each task, we can conclude that these are really promising results. Additionally, the FoodNER models trained in the tasks of distinguishing food entities concerning semantic tags on the level of food groups are the first corpus-based NERs that can distinguish between different food semantic tags (ie, food groups). Once more, we should emphasize that in the cases of FoodOn and SNOMED CT, the BERT and BioBERT models are tuned only on the entities that have semantic tags provided by the FoodOntoMap resource, in which not all food entities from the semantic resources are present.

Table 2. Macro F1 scores for the 3 food named-entity recognition models for the tasks concerning different semantic models.

Model, semantic model	Epochs[a]	Macro F1 score (%)
BERT[b]		
FoodOn	100	78.13
Hansard closest	85	75.87
Hansard parent	100	75.04
SNOMED CT[c]	91	76.01
BioBERT-large		
FoodOn	93	75.58
Hansard closest	100	78.96
Hansard parent	100	76.26
SNOMED CT	95	74.51
BioBERT-standard		
FoodOn	100	74.81
Hansard closest	100	74.18
Hansard parent	89	74.94
SNOMED CT	89	73.39

[a] This provides information on the number of epochs needed to fine-tune the model.

[b] BERT: bidirectional encoder representations from transformers.

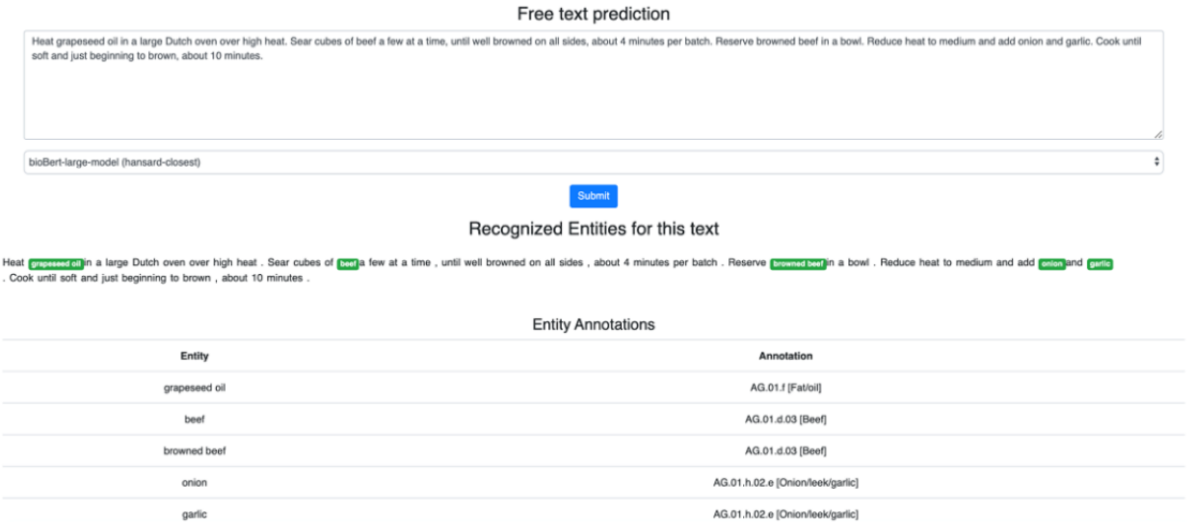
[c] SNOMED CT: Systematized Nomenclature of Medicine Clinical Terms.

Discussion

Principal Findings

The models are trained on FoodBase [38], in which recipes that are collected from the biggest social media networks for sharing and discovering recipes, are annotated. Since this is a specific type of text, there are some weaknesses when it comes to applications on texts of a different nature (eg, medical texts). To address this in our future work, we plan to further retrain the models on various types of documents such as dietary recommendations and PubMed articles. Regardless of this, the presented BERT models are robust for extracting food concepts while simultaneously normalizing them to some semantic resource, which allows further interlinking of the entities with other domains (eg, health and environmental sciences). This will help to improve the quality of health and clinical practices. The semantic tags were selected based on the food annotations that exist from the FoodBase and FoodOntoMap resources. However, in future, the FoodNER methodology may be applied on any other annotated corpus from this domain. To bring our work closer to subject matter experts from the food domain, the FoodNER models have been integrated in the FoodViz platform [45]. Figure 7 shows the interface where subject matter experts can place an arbitrary recipe, select a model, and preview the annotated food entities. We provide highlighting of the phrases in the text, as well as the tabular display of the food phrases and their annotations. Figure 7 is an example where a short description from a recipe *“Heat rapeseed oil in a large Dutch oven over high heat. Sear cubes of beef a few at a time, until well browned on all sides, about 4 minutes per batch. Reserve browned beef in a bowl. Reduce heat to medium and add onion and garlic. Cook until soft and just beginning to brown, about 10 minutes.”* is annotated using the model fine-tuned with BioBERT large in the Hansard closest task. From the annotations provided, it is obvious that the model can recognize all food entities that are mentioned in the text (ie, *grapeseed oil*, *beef*, *browned beef*, *onion*, and *garlic*) annotated by Hansard semantic tags. This interface radically simplifies the usage of the state-of-the-art models for subject matter experts in the food domain, without their knowledge of the underlying details, such as machine learning or IOB format understanding. Additionally, the current architecture of the FoodViz application allows integration of new prediction models only with their upload at the corresponding location in the server.

Figure 7. Food named-entity recognition integration in FoodViz.



Free text prediction

Heat grapeseed oil in a large Dutch oven over high heat. Sear cubes of beef a few at a time, until well browned on all sides, about 4 minutes per batch. Reserve browned beef in a bowl. Reduce heat to medium and add onion and garlic. Cook until soft and just beginning to brown, about 10 minutes.

bioBert-large-model (hansard-closest)

Recognized Entities for this text

Heat grapeseed oil in a large Dutch oven over high heat . Sear cubes of beef a few at a time , until well browned on all sides , about 4 minutes per batch . Reserve browned beef in a bowl . Reduce heat to medium and add onion and garlic . Cook until soft and just beginning to brown , about 10 minutes .

Entity Annotations

Entity	Annotation
grapeseed oil	AG.01.f [Fat/oil]
beef	AG.01.d.03 [Beef]
browned beef	AG.01.d.03 [Beef]
onion	AG.01.h.02.e [Onion/leek/garlic]
garlic	AG.01.h.02.e [Onion/leek/garlic]

Conclusion

We present a corpus-based NER method for food information extraction, known as FoodNER. It is developed by fine-tuning the BERT model by using 3 previously published predefined BERT representation language models (ie, the original BERT and 2 BioBERTs; standard and large). FoodNER can be used to extract and annotate food entities in 5 different tasks: distinguishing between food versus nonfood entities and distinguishing food entities on the level of food groups by using the closest Hansard semantic tags, the parent Hansard semantic tags, the FoodOn semantic tags, or the SNOMED CT semantic tags. All in all, the models provide very promising results achieving around 93.30%-94.31% macro F1 scores in the food versus nonfood entity task and around 73.39%-78.96% macro F1 scores in the tasks where more semantic tags are recognized. Additionally, the models are included in the FoodViz framework, which allows users to select which FoodNER model they want to use for the annotation of their texts with food entities and additionally provides a visualization of the annotated data with an opportunity to correct the false positive and false negative annotations. Having such a robust state-of-the-art food information extraction method such as FoodNER will allow further research in investigating food-drug and food-disease interactions, thereby providing an opportunity to start building a food knowledge graph, including relations with health-related entities.

Acknowledgments

This research was supported by the Slovenian Research Agency (research core grant P2-0098 and grant PR-10465), the European Union’s Horizon 2020 research and innovation program (FNS-Cloud, Food Nutrition Security) (grant agreement 863059), and the Ad Futura grant for postgraduate study. The information and the views set out in this publication are those of the authors and do not necessarily reflect the official opinion of the European Union. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use that may be made of the information contained herein.

Conflicts of Interest

None declared.

References

1. Johan F, Owen G. Scaling 36 solutions to halve emissions by 2030. Exponential Roadmap. 2020. URL: https://exponentialroadmap.org/wp-content/uploads/2019/09/Exponential-Roadmap-1.5-September-19-2019.pdf [accessed 2021-05-19]

2. Qiao L, Yang L, Hong D, Yao L, Zhiguang Q. Knowledge graph construction techniques. Journal of computer research and development 2016;53(3):582 [FREE Full text] [doi: 10.7544/issn1000-1239.2016.20148228]
3. Zhou X, Zhang X, Hu X. MaxMatcher: Biological concept extraction using approximate dictionary lookup. 2006 Presented at: Pacific Rim International Conference On Artificial Intelligence; August 7-11, 2006; Guilin, China p. 1145-1149. [doi: 10.1007/11801603_150]
4. Hanisch D, Fundel K, Mevissen H, Zimmer R, Fluck J. ProMiner: rule-based protein and gene entity recognition. BMC Bioinformatics 2005;6(Suppl 1):S14. [doi: 10.1186/1471-2105-6-s1-s14]
5. Eftimov T, Koroušić Seljak B, Korošec P. A rule-based named-entity recognition method for knowledge extraction of evidence-based dietary recommendations. PLoS One 2017;12(6):e0179488 [FREE Full text] [doi: 10.1371/journal.pone.0179488] [Medline: 28644863]
6. Alnazzawi N, Thompson P, Batista-Navarro R, Ananiadou S. Using text mining techniques to extract phenotypic information from the PhenoCHF corpus. BMC Med Inform Decis Mak 2015 Jun 15;15(S2):1-10. [doi: 10.1186/1472-6947-15-s2-s3]
7. Leaman R, Wei CH, Zou C, Lu Z. Mining patents with tmChem, GNormPlus and an ensemble of open systems. Proceedings of the Fifth Biocreative Challenge Evaluation Workshop. URL: https://biocreative.bioinformatics.udel.edu/media/store/files/2015/BCV2015_paper_61.pdf [accessed 2021-07-05]
8. Settles B. Active learning literature survey. University of Wisconsin Madison. URL: https://minds.wisconsin.edu/bitstream/handle/1793/60660/TR1648.pdf?sequence=1&isAllowed=y [accessed 2021-05-19]
9. Lopez MM, Kalita J. Deep learning applied to NLP. Cornell University. URL: https://arxiv.org/abs/1703.03091 [accessed 2021-06-24]
10. Devlin J, Chang MW, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. Cornell University. URL: https://arxiv.org/abs/1810.04805 [accessed 2021-06-24]
11. Dang TH, Le HQ, Nguyen TM, Vu ST. D3NER: biomedical named entity recognition using CRF-biLSTM improved with fine-tuned embeddings of various linguistic information. Bioinformatics 2018 Oct 15;34(20):3539-3546. [doi: 10.1093/bioinformatics/bty356] [Medline: 29718118]
12. Giorgi JM, Bader GD. Transfer learning for biomedical named entity recognition with neural networks. Bioinformatics 2018 Dec 01;34(23):4087-4094 [FREE Full text] [doi: 10.1093/bioinformatics/bty449] [Medline: 29868832]
13. Yoon W, So CH, Lee J, Kang J. CollaboNet: collaboration of deep neural networks for biomedical named entity recognition. BMC Bioinformatics 2019 May 29;20(Suppl 10):249 [FREE Full text] [doi: 10.1186/s12859-019-2813-6] [Medline: 31138109]
14. Wang X, Zhang Y, Ren X, Zhang Y, Zitnik M, Shang J, et al. Cross-type biomedical named entity recognition with deep multi-task learning. Bioinformatics 2019 May 15;35(10):1745-1752. [doi: 10.1093/bioinformatics/bty869] [Medline: 30307536]
15. Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics 2020 Feb 15;36(4):1234-1240 [FREE Full text] [doi: 10.1093/bioinformatics/btz682] [Medline: 31501885]
16. Zhou H, Ning S, Liu Z, Lang C, Liu Z, Lei B. Knowledge-enhanced biomedical named entity recognition and normalization: application to proteins and genes. BMC Bioinformatics 2020 Jan 30;21(1):35 [FREE Full text] [doi: 10.1186/s12859-020-3375-3] [Medline: 32000677]
17. Cho M, Ha J, Park C, Park S. Combinatorial feature embedding based on CNN and LSTM for biomedical named entity recognition. J Biomed Inform 2020 Mar;103:103381 [FREE Full text] [doi: 10.1016/j.jbi.2020.103381] [Medline: 32004641]
18. Lindberg DAB, Humphreys BL, McCray AT. The Unified Medical Language System. Yearb Med Inform 2018 Mar 05;02(01):41-51. [doi: 10.1055/s-0038-1637976]
19. Arighi CN, Carterette B, Cohen KB, Krallinger M, Wilbur WJ, Fey P, et al. An overview of the BioCreative 2012 Workshop Track III: interactive text mining task. Database (Oxford) 2013;2013:bas056 [FREE Full text] [doi: 10.1093/database/bas056] [Medline: 23327936]
20. Mao Y, Van Auken K, Li D, Arighi CN, McQuilton P, Hayman GT, et al. Overview of the gene ontology task at BioCreative IV. Database (Oxford) 2014;2014:bau086 [FREE Full text] [doi: 10.1093/database/bau086] [Medline: 25157073]
21. Wei CH, Peng Y, Leaman R, Davis AP, Mattingly CJ, Li J, et al. Assessing the state of the art in biomedical relation extraction: overview of the BioCreative V chemical disease relation (CDR) task. Database (Oxford) 2016;2016:baw032. [doi: https:10.1093/database/baw032]
22. Pyysalo S, Ohta T, Rak R, Rowley A, Chun H, Jung S, et al. Overview of the Cancer Genetics and Pathway Curation tasks of BioNLP Shared Task 2013. BMC Bioinformatics 2015 Jun 23;16(S10):1-19. [doi: 10.1186/1471-2105-16-s10-s2]
23. Stubbs A, Kotfila C, Uzuner. Automated systems for the de-identification of longitudinal clinical narratives: Overview of 2014 i2b2/UTHealth shared task Track 1. J Biomed Inform 2015 Dec;58 Suppl:S11-S19 [FREE Full text] [doi: 10.1016/j.jbi.2015.06.007] [Medline: 26225918]
24. Deléger L, Bossy R, Chaix E, Ba M, Ferre A, Bessieres P, et al. Overview of the bacteria biotope task at bioNLP shared task 2016. 2016 Presented at: Proceedings of the 4th BioNLP shared task workshop; August 12, 2016; Berlin, Germany p. 12-22. [doi: 10.18653/v1/w16-3002]

25. Cohen KB, Demner-Fushman D, Ananiadou S, Tsujii J. Biomedical natural language processing in 2017: The view from computational linguistics. The Association for Computational Linguistics. URL: https://aclanthology.org/W17-23.pdf [accessed 2021-06-24]
26. Wang Y, Zhou K, Gachloo M, Xia J. An Overview of the Active Gene Annotation Corpus and the BioNLP OST 2019 AGAC Track Tasks. 2019 Presented at: Proceedings of The 5th Workshop on BioNLP Open Shared Tasks; November 3-7, 2019; Hong Kong p. 62-71. [doi: 10.18653/v1/d19-5710]
27. Ernst P, Siu A, Weikum G. KnowLife: a versatile approach for constructing a large knowledge graph for biomedical sciences. BMC Bioinformatics 2015 May 14;16:157 [FREE Full text] [doi: 10.1186/s12859-015-0549-5] [Medline: 25971816]
28. Sang S, Yang Z, Liu X, Wang L, Lin H, Wang J, et al. GrEDeL: A Knowledge Graph Embedding Based Method for Drug Discovery From Biomedical Literatures. IEEE Access 2019;7:8404-8415. [doi: 10.1109/access.2018.2886311]
29. Boulos M, Yassine A, Shirmohammadi S, Namahoot C, Brückner M. Towards an “Internet of Food”: Food Ontologies for the Internet of Things. Future Internet 2015 Oct 01;7(4):372-392. [doi: 10.3390/fi7040372]
30. Yang C, Ambayo H, Baets BD, Kolsteren P, Thanintorn N, Hawwash D, et al. An Ontology to Standardize Research Output of Nutritional Epidemiology: From Paper-Based Standards to Linked Content. Nutrients 2019 Jun 08;11(6):1300 [FREE Full text] [doi: 10.3390/nu11061300] [Medline: 31181762]
31. Popovski G, Koroušić Seljak B, Eftimov T. FoodIE: A Rule-based Named-entity Recognition Method for Food Information Extraction. 2019 Presented at: International Conference on Pattern Recognition Applications and Methods; February 19-21, 2019; Prague, Czech Republic p. 915-922. [doi: 10.5220/0007686309150922]
32. Alexander M, Anderson J. The Hansard corpus, 1803-2003. University of Glasgow. URL: https://eprints.gla.ac.uk/81804/ [accessed 2021-05-19]
33. Jonquet C, Shah N, Youn C, Callendar C, Storey MA, Musen M. NCBO annotator: semantic annotation of biomedical data. International Semantic Web Conference. URL: http://www.lirmm.fr/~jonquet/publications/documents/Demo-ISWC09-Jonquet.pdf [accessed 2021-06-24]
34. Noy NF, Shah NH, Whetzel PL, Dai B, Dorf M, Griffith N, et al. BioPortal: ontologies and integrated data resources at the click of a mouse. Nucleic Acids Res 2009 Jul;37(Web Server issue):W170-W173 [FREE Full text] [doi: 10.1093/nar/gkp440] [Medline: 19483092]
35. Dooley DM, Griffiths EJ, Gosal GS, Buttigieg PL, Hoehndorf R, Lange MC, et al. FoodOn: a harmonized food ontology to increase global food traceability, quality control and data integration. NPJ Sci Food 2018;2:23 [FREE Full text] [doi: 10.1038/s41538-018-0032-6] [Medline: 31304272]
36. Donnelly K. SNOMED-CT: The advanced terminology and coding system for eHealth. Stud Health Technol Inform 2006;121:279-290. [Medline: 17095826]
37. Popovski G, Seljak B, Eftimov T. A Survey of Named-Entity Recognition Methods for Food Information Extraction. IEEE Access 2020;8:31586-31594. [doi: 10.1109/access.2020.2973502]
38. Popovski G, Koroušić Seljak B, Eftimov T. FoodBase corpus: a new resource of annotated food entities. Database 2019:baz121. [doi: 10.1093/database/baz121]
39. Comeau DC, Islamaj Doğan R, Ciccarese P, Cohen KB, Krallinger M, Leitner F, et al. BioC: a minimalist approach to interoperability for biomedical text processing. Database (Oxford) 2013;2013:bat064 [FREE Full text] [doi: 10.1093/database/bat064] [Medline: 24048470]
40. Cenikj G, Popovski G, Stojanov R, Koroušić Seljak B, Eftimov T. BuTTER: BidirecTional LSTM for Food Named-Entity Recognition. 2020 Presented at: 2020 IEEE International Conference on Big Data (Big Data); December 10-13, 2020; Virtual p. 3550-3556. [doi: 10.1109/bigdata50022.2020.9378151]
41. Popovski G, Koroušić Seljak B, Eftimov T. FoodOntoMap: Linking Food Concepts across Different Food Ontologies. 2019 Presented at: 11th International Conference on Knowledge Engineering and Ontology Development; 2019; Vienna, Austria p. 195-202. [doi: 10.5220/0008353201950202]
42. Akhtyamova L. Named entity recognition in Spanish biomedical literature: Short review and BERT model. 2020 Presented at: 26th Conference of Open Innovations Association (FRUCT); April 20-24, 2020; Yaroslavl, Russia p. 1-7. [doi: 10.23919/fruct48808.2020.9087359]
43. Kim J, Ko Y, Seo J. Construction of Machine-Labeled Data for Improving Named Entity Recognition by Transfer Learning. IEEE Access 2020;8:59684-59693. [doi: 10.1109/access.2020.2981361]
44. Miftahutdinov Z, Alimova I, Tutubalina E. On biomedical named entity recognition: experiments in interlingual transfer for clinical and social media texts. 2020 Presented at: European Conference on Information Retrieval; April 14-17, 2020; Lisbon, Portugal p. 281-288. [doi: 10.1007/978-3-030-45442-5_35]
45. Stojanov R, Popovski G, Jofce N, Dimitar T, Koroušić Seljak B, Eftimov T. Foodviz: Visualization of food entities linked across different standards. 2020 Presented at: International Conference on Machine Learning, Optimization, and Data Science; July 19-23, 2020; Siena-Tuscany, Italy p. 28-38. [doi: 10.1007/978-3-030-64580-9_4]
46. Rayson P, Archer D, Piao S, McEnery AM. The UCREL semantic analysis system. 2004 Presented at: Proceedings of the beyond named entity recognition semantic labelling for NLP tasks workshop; January 2004; Lisbon, Portugal p. 7-12 URL: https://www.researchgate.net/publication/228881331_The_UCREL_semantic_analysis_system

47. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet Large Scale Visual Recognition Challenge. Int J Comput Vis 2015 Apr 11;115(3):211-252. [doi: 10.1007/s11263-015-0816-y]
48. Conneau A, Kiela D, Schwenk H, Barrault L, Bordes A. Supervised learning of universal sentence representations from natural language inference data. ACL Anthology 2017:1-12. [doi: 10.18653/v1/d17-1070]
49. Zhu Y, Kiros R, Zemel R, Salakhutdinov R, Urtasun R, Torralba A, et al. Aligning books and movies: Towards story-like visual explanations by watching movies and reading books. 2015 Presented at: Proceedings of the IEEE international conference on computer vision; December 7-13, 2015; Santiago, Chile p. 19-27. [doi: 10.1109/iccv.2015.11]
50. Ramshaw LA, Marcus MP. Text chunking using transformation-based learning. Natural language processing using very large corpora 1999:157-176. [doi: 10.1007/978-94-017-2390-9_10]
51. Hung Z, Xu W, Yu K. Bidirectional LSTM-CRF models for sequence tagging. Cornell University. 2015. URL: https://arxiv.org/abs/1508.01991 [accessed 2021-07-05]
52. Pennington J, Socher R, Manning CD. GloVe: Global vectors for word representation. 2014 Presented at: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP); October 25-29, 2014; Doha, Qatar p. 1532-1543. [doi: 10.3115/v1/d14-1162]
53. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. Cornell University. URL: https://arxiv.org/abs/1301.3781 [accessed 2021-07-05]
54. Stojanov R, Popovski G, Cenikj G, Koroušić Seljak B, Eftimov T. The source code (IPython notebook). GitHub. URL: https://github.com/ds4food/FoodNer/blob/master/FoodNER.ipynb [accessed 2021-05-19]
55. Bojanowski P, Grave E, Joulin A, Mikolov T. Enriching Word Vectors with Subword Information. TACL 2017 Dec;5:135-146. [doi: 10.1162/tacl_a_00051]
56. Stojanov R, Popovski G, Cenikj G, Koroušić Seljak B, Eftimov T. Detailed statistics about each of the fine-tuned models. FoodViz with FoodNER. URL: http://foodviz.env4health.finki.ukim.mk/#/result-resources [accessed 2021-05-19]
57. Stojanov R, Popovski G, Cenikj G, Koroušić Seljak B, Eftimov T. FoodNER. GitHub. URL: https://github.com/ds4food/FoodNer/tree/master/data [accessed 2021-05-19]

Abbreviations

BERT: bidirectional encoder representations from transformers
BiLSTM-CRF: bidirectional long short-term memory conditional random field
BuTTER: bidirectional long short-term memory for food named-entity recognition
IOB: inside, outside, and beginning
NCBO: National Center for Biomedical Ontology
NER: named-entity recognition
NLP: natural language processing
SNOMED CT: Systematized Nomenclature of Medicine Clinical Terms

---

Edited by R Kukafka; submitted 25.02.21; peer-reviewed by C Yang, M Torii; comments to author 08.03.21; revised version received 13.03.21; accepted 06.05.21; published 09.08.21.

Please cite as:
Stojanov R, Popovski G, Cenikj G, Koroušić Seljak B, Eftimov T
A Fine-Tuned Bidirectional Encoder Representations From Transformers Model for Food Named-Entity Recognition: Algorithm Development and Validation
J Med Internet Res 2021;23(8):e28229
URL: https://www.jmir.org/2021/8/e28229
doi:10.2196/28229
PMID:34383671

©Riste Stojanov, Gorjan Popovski, Gjorgjina Cenikj, Barbara Koroušić Seljak, Tome Eftimov. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 09.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Automated Screening for Social Anxiety, Generalized Anxiety, and Depression From Objective Smartphone-Collected Data: Cross-sectional Study

Daniel Di Matteo[1], PhD; Kathryn Fotinos[2], BSc (Hons); Sachinthya Lokuge[2], BSc (Hons); Geneva Mason[2], BASc (Hons); Tia Sternat[2,3], MPsy; Martin A Katzman[2,3,4,5], MD; Jonathan Rose[1], PhD, FIEEE, FACM, FM NAE, FRSC

[1]The Edward S Rogers Sr Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada

[2]START Clinic for Mood and Anxiety Disorders, Toronto, ON, Canada

[3]Department of Psychology, Adler Graduate Professional School, Toronto, ON, Canada

[4]Department of Psychology, Lakehead University, Thunder Bay, ON, Canada

[5]The Northern Ontario School of Medicine, Thunder Bay, ON, Canada

Corresponding Author:
Daniel Di Matteo, PhD
The Edward S Rogers Sr Department of Electrical and Computer Engineering
University of Toronto
10 King's College Road
Toronto, ON
Canada
Phone: 1 416 978 6992
Email: daniel.dimatteo@utoronto.ca

Abstract

Background: The lack of access to mental health care could be addressed, in part, through the development of automated screening technologies for detecting the most common mental health disorders without the direct involvement of clinicians. Objective smartphone-collected data may contain sufficient information about individuals’ behaviors to infer their mental states and therefore screen for anxiety disorders and depression.

Objective: The objective of this study is to compare how a single set of recognized and novel features, extracted from smartphone-collected data, can be used for predicting generalized anxiety disorder (GAD), social anxiety disorder (SAD), and depression.

Methods: An Android app was designed, together with a centralized server system, to collect periodic measurements of objective smartphone data. The types of data included samples of ambient audio, GPS location, screen state, and light sensor data. Subjects were recruited into a 2-week observational study in which the app was run on their personal smartphones. The subjects also completed self-report severity measures of SAD, GAD, and depression. The participants were 112 Canadian adults from a nonclinical population. High-level features were extracted from the data of 84 participants, and predictive models of SAD, GAD, and depression were built and evaluated.

Results: Models of SAD and depression achieved a significantly greater screening accuracy than uninformative models (area under the receiver operating characteristic means of 0.64, SD 0.13 and 0.72, SD 0.12, respectively), whereas models of GAD failed to be predictive. Investigation of the model coefficients revealed key features that were predictive of SAD and depression.

Conclusions: We demonstrate the ability of a common set of features to act as predictors in the models of both SAD and depression. This suggests that the types of behaviors that can be inferred from smartphone-collected data are broad indicators of mental health, which can be used to study, assess, and track psychopathology simultaneously across multiple disorders and diagnostic boundaries.

(J Med Internet Res 2021;23(8):e28918) doi:10.2196/28918

KEYWORDS

mobile sensing; passive EMA; passive sensing; psychiatric assessment; mood and anxiety disorders; mobile apps; mhealth; mobile phone; digital health; digital phenotyping

Introduction

Background

Problems related to mental health are widespread, with 1 report estimating that 1 in 5 Canadians experience a problem related to mental health or addiction in any given year [1]. In 2018, 2.3 million Canadians went without adequate support for their mental health, with often-cited reasons including lack of information, time, or finances [2]. These impediments could potentially be addressed with the development and distribution of technology to aid in the diagnosis, tracking, and treatment of mental health disorders.

The widespread use of smartphones coupled with their sensing capabilities and persistent network connectivity makes them a promising platform for delivering these technologies. The sensors present on all modern smartphones offer a rich landscape of data that can be used to infer behavior or actions pertinent to an individual's mental health state. These data could then serve as the basis for screening algorithms for mood and anxiety disorders such as depression, generalized anxiety disorder (GAD), and social anxiety disorder (SAD). The development of automated screening may be a crucial step in mental health care. Automated screeners could offer a low-cost monitoring system to individuals struggling with mental health problems and provide potentially useful information for practitioners of mental health care.

Previous Work

A core aspect of this work is the extraction of relevant knowledge from the wide array of smartphone-sensed data. To convey a sense of scale of this class of data, consider that smartphones can sense, measure, and record geolocation, motion, ambient lighting conditions, ambient audio, communication patterns, and app use in real time. In a matter of weeks, tens of thousands of data points across these many data modalities can easily be collected from a single individual. Nonetheless, even with all these data, it remains quite challenging to build predictive models that are both accurate *and* transparent, in that they offer insight into what specific patterns of data, and what points in time, are driving models to make specific decisions or characterizations [3]. It is our goal with this work to address this challenge by building predictive models that are both accurate and transparent.

A general approach to this problem of knowledge extraction and interpretability has been to review the diagnostic criteria and the key symptoms of a mental health disorder and to look for data streams that would relate to these. In this approach, data are transformed into relevant knowledge by determining if the presence or absence of a symptom or characteristic of a disorder can be inferred from the available data. For example, key symptoms and criteria for depression include sleep disturbances (insomnia or hypersomnia) [4] and low energy [5]. The duration of sleep may be inferred by analyzing a combination of light and motion sensor data, screen activity, and ambient audio [6]. Although a person's energy level is difficult to measure directly, proxy measures can be effectively created by observing the degree to which subjects leave their homes and travel throughout their environments [7] or by observing activity levels as measured by motion sensors [8].

This approach of using characteristics of disorder symptomology to drive the understanding of smartphone-collected data has been used in the modeling of SAD. Individuals with social anxiety are characterized by fear or avoidance of social interactions, ultimately stemming from the fear of negative evaluation from others [9]. Although the fear of evaluation is most certainly a latent construct, the avoidance of social situations can be objectively measured from smartphone-sensed data. One example of how this can be achieved is through the classification of different locations as inherently *social* or not, achievable using so-called *semantic location* data. Measuring the degree to which subjects travel to these inherently social locations (such as pubs and dancing venues) has been shown to be correlated with social anxiety levels [10]. Chow et al [11] demonstrated that greater time spent at home was correlated with stronger symptoms of social anxiety.

At the time of this writing, the authors are unaware of any studies that have used passively collected smartphone data to infer behaviors associated with GAD or to otherwise predict or screen for GAD.

With respect to depression, Ben-Zeev et al [12] demonstrated associations between the severity of depressive symptoms and smartphone-sensed geospatial activity, speech duration, and sleep duration. In a study that focused solely on GPS location–based data, Canzian and Musolesi [13] found associations between distance traveled and depressive symptoms. Increased time spent at home, as inferred from smartphone-collected GPS data, was also found to be associated with stronger symptoms of depression in studies by Saeb et al [7] and Farhan et al [14]. Both studies also found that a greater number of unique locations visited by subjects was associated with weaker symptoms of depression [7,14]. Studies by Wang et al [15] and Ben-Zeev et al [12] also used smartphone-recorded audio data to build a proxy measure for social interaction and found that this measure was negatively associated with depressive symptoms. Readers interested in a broader review of the associations between symptoms of mood disorders (including depression and additionally bipolar disorder) and smartphone-collected data are requested to refer the excellent review article by Rohani et al [16].

Goal of This Study

This study aims to expand upon the existing body of work in using mobile sensing to predict certain aspects of mental health. Although the body of knowledge in this field continues to grow, much of it mainly focuses on mood disorders (depression and bipolar disorder), with much less of this work being undertaken in anxiety disorders. Studies have shown key features that act as predictors of depression, but many of these features have been motivated by general trends or associations with mental health that may extend beyond mood disorders to anxiety disorders. Furthermore, because anxiety disorders are commonly comorbid with depression [17], features that have been successful in predicting depression may also be successful in predicting anxiety disorders. The goal of this study is to propose a set of features, some novel and some existing, that have already been shown to be successful predictors of depression and to use this set of features in the independent prediction of GAD, SAD, and depression.

Methods

Recruitment and Demographics

The subjects were adults recruited from the general population using Prolific, a web-based recruitment platform [18]. Recruitment was conducted from July 2019 to December 2019. Subjects were not prescreened for psychiatric diagnoses. The study inclusion criteria were as follows: subjects must (1) reside in Canada, (2) be fluent in English, (3) own an Android phone, (4) have completed at least 95% of their previous Prolific studies successfully, and (5) have previously participated in at least 20 Prolific studies. The final criterion was used to ensure that subjects were proficient in using the Prolific system and were generally technology literate. There were no exclusion criteria for this study. Subjects were paid Can $18.50 (US $14.20) to participate in the study.

Study Procedure

Subjects were entered into a 14-day observational study where a custom Android app was installed on their personal smartphone. The app collected both self-report measures of mental health (collected at the beginning and end of the study) and objective smartphone-sensed data, collected periodically and completely passively throughout the entire 2 weeks of the study.

Self-report Measures and Mental Health Screening

Subjects completed 4 self-report measures, in digital form, within the study app, at the beginning and end of the 14-day study. A review by Belisario et al [19] found that self-administered survey scores did not differ significantly when deployed by the app compared with other delivery modes [19]. These self-reported surveys were completed on their own with no supervision by clinicians. Subjects completed the following four self-report measures of mental health: (1) the Liebowitz Social Anxiety Scale (LSAS), which is a 24-item self-report scale used in the assessment of SAD [20]; (2) the Generalized Anxiety Disorder 7-item (GAD-7) scale, which is a screening and assessment tool for GAD [21]; (3) the Patient Health Questionnaire 8-item (PHQ-8) scale, which is a screener and assessment tool for depression [22]; and (4) the Sheehan Disability Scale, which is a 3-item scale that assesses general impairment due to mental health [23]. Data from the Sheehan Disability Scale were not used in any of the analyses in this study.

Although all scales were completed at both the study intake and exit, only the values obtained at the exit were used in the analysis presented in this paper. This was done because the self-report measures asked respondents to evaluate symptoms over the *previous* 2 weeks; therefore, the window of symptom self-assessment at the study exit coincides with the window of electronic data collection. Changes with respect to baseline (ie, differences between scores measured at intake and exit) were not investigated because the goal was solely to determine whether digital technology could perform automated screenings from short-term data collection.

Exit values on the LSAS, GAD-7, and PHQ-8 were used to screen for SAD, GAD, and major depressive disorder (MDD), respectively, by applying thresholds to scores. A threshold of 60 was used with the LSAS scores to screen for SAD (generalized subtype), as recommended by Mennin et al [24]. This threshold was shown to screen for generalized social anxiety with sensitivity of 82% and specificity of 78% [25]. A threshold of 10 was used with the GAD-7 scores to screen for GAD. This threshold was shown to optimize sensitivity (89%) and specificity (82%) [21]. A threshold of 10 was used with the PHQ-8 scores to screen for depression, as recommended by Kroenke et al [22]. This threshold was shown to screen for major depression with sensitivity of 88% and specificity of 88% [22].

Smartphone-Sensed Data Collection

The app collected data across 4 streams of data available on an Android smartphone: audio (from the microphone), geolocation (GPS sensor), screen state (screen on or off), and illuminance of the environment (light sensor).

Audio data were collected by using devices' microphones to record the ambient audio of the environment for a 15-second duration, at a nominal rate of one 15-second recording every 5 minutes. To preserve privacy, these audio recordings were processed to yield 3 less invasive streams of data, after which the recordings were automatically destroyed. The first stream of audio-derived data is simply the average volume of each 15-second audio recording. The second stream of audio-derived data is a label for each audio recording, which indicates whether English-language speech was detected in the environmental audio at the time of each recording. Automatic speech recognition (ASR) software [26] was used to generate transcripts from each audio recording, and recordings that produced empty transcripts were labeled as containing no speech, whereas recordings that produced transcripts were labeled as containing speech. Finally, the third stream of audio-derived data is the entire set of ASR-detected English words. These words were stored in a randomized order, with no associated time information, to prevent the recreation of the original transcripts. Audio recordings were deleted by the software immediately after the extraction of mean volume, speech presence labels, and detected words.

The GPS location (latitude and longitude) was recorded at a nominal rate of once every 5 minutes. The state of the screen was not sampled periodically, but instead, the app recorded screen transitions (ie, turned off or on) whenever the state of the screen changed. Finally, the illuminance of the subjects' environments, as measured by the devices' light sensors, was recorded at a nominal rate of once every 10 minutes. We refer to these data sampling rates as nominal because, in practice, the observed frequencies are often lower than what was specified in the software because of the battery-preserving restriction imposed by the Android operating system in devices running Android version 6 and above [27].

Feature Extraction

Feature Extraction Overview

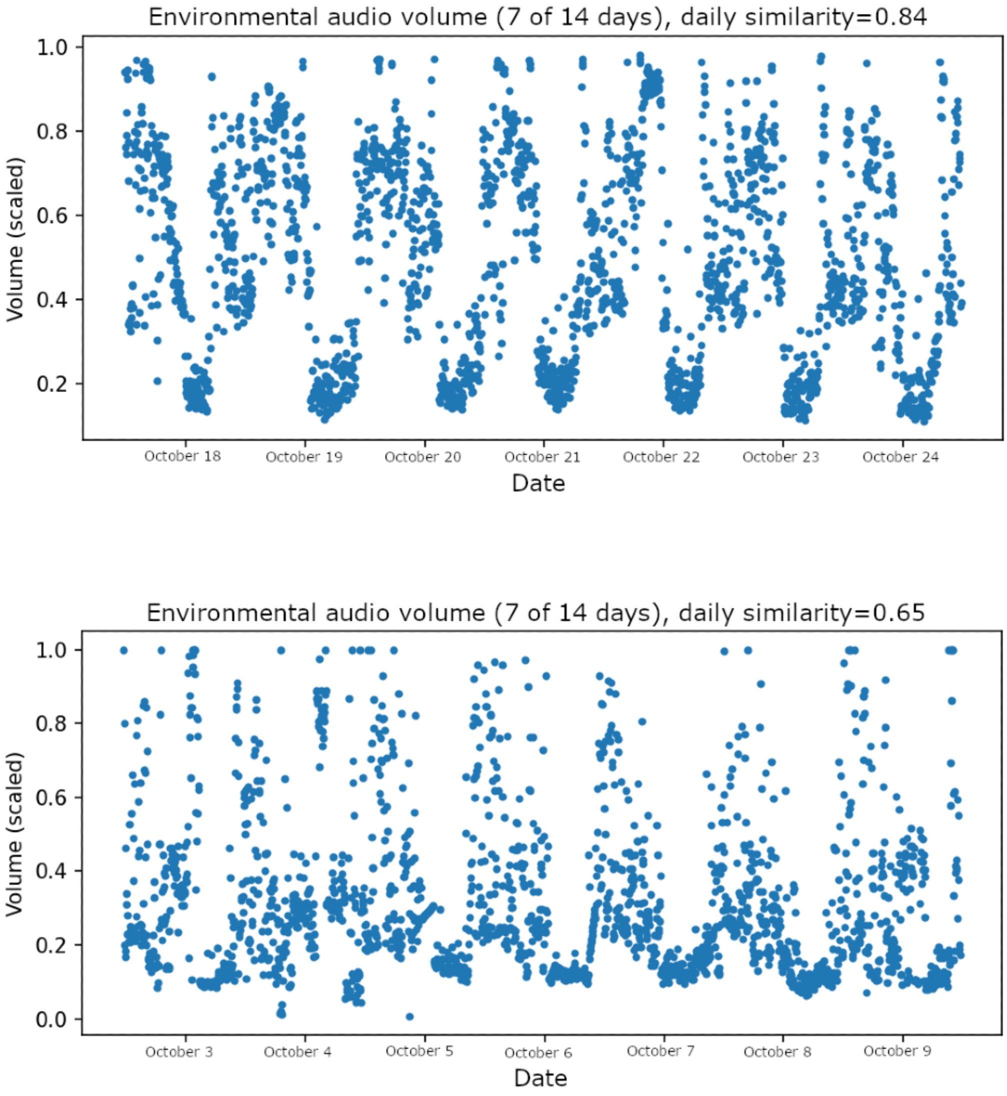
Predictor variables, or *features*, were extracted from the raw smartphone-collected sensor data to act as inputs to the predictive models of SAD, GAD, and depression. These features are all hypothesized as capturing behaviors that are relevant to mental state, with the goal of using them as explanatory variables within the prediction models. The following sections describe each of the features that were used.

Daily Similarity

The ability to establish and maintain regularity in one's patterns of activities has been associated with positive mental health [28,29]. Assuming that the volume of subjects' environments could be interpreted as a proxy for activity, we developed a feature, called *daily similarity*, which quantifies how periodic the volume of the audio recordings was, with respect to a 24-hour or 1-day period. When the volume of the audio recordings was extracted and treated as time series data, there were clear peaks and troughs in the signal, corresponding to the daytime activity and nighttime inactivity, respectively. It was hypothesized that individuals with more regular patterns of activity and inactivity, as inferred from the peaks and troughs of this time series data, would have better mental health. The daily similarity feature was designed to compute a quantitative measure of the regularity, in time, of the sequence of these peaks and troughs—in other words, to what degree peaks and troughs occur at the same time across days.

Mathematically, daily similarity is computed as the autocorrelation function of the volume time series of ambient audio recordings evaluated at a lag time of 24 hours. Figure 1 provides an illustration of 2 study participants' volume time series data, with the computed daily similarity feature also annotated. Note that the first participant's data, presented in the top half of the figure, is slightly more regular and has a higher daily similarity value than the participant whose data are presented in the bottom half of the figure. This feature was first presented and analyzed in a previous study by Di Matteo et al [30]. Note that the daily similarity feature, as computed in this work, only captures similarity on a short time scale; as a result, it is not clear if long-term stability in social rhythms would be captured.

Figure 1. Environmental audio volume timeseries and values of the daily similarity feature for 2 study participants.



Speech Presence

The avoidance of social interaction is a defining characteristic of SAD [31] and is also evident among individuals with depression [32]. The *speech presence* feature serves as a measure of social interaction by quantifying the presence of speaking voices in the periodic audio recordings. The speech presence feature is defined as the number of audio recordings labeled as containing speech divided by the total number of audio recordings. This feature was first proposed by Wang et al [15] and was confirmed to be a good correlate of anxiety and depression in our previous study [30]. We note that *any* speech in the environment, including the playback of recorded speech, will be detected as speech; therefore, this feature is not a perfect proxy for social interaction. The results of the study by Di Matteo et al [30] were reported under the same conditions.

Weeknight Sleep Disturbance

Sleep is also known to be a key factor in mental health disorders, with sleep disturbances being a common feature of anxiety and depression [4,33]. We compute a feature called *weeknight sleep disturbance*, which infers a proxy measure of sleep quality from the same audio volume time series data as the daily similarity feature. The underlying assumption of this feature is that when a subject's environment is noisy during common hours of sleep, then it is likely that that the subject is either awake or not sleeping well. Specifically, this feature is a measure of the noisiness of a subject's environment during common sleeping hours. The weeknight sleep disturbance feature is computed as the SD of the volume time series of ambient audio recordings for weeknights between midnight and 6 AM. Weekend evenings are excluded because some individuals may shift their sleep schedules on weekends, if they work weekdays and not weekends. The SD of the volume time series is used as a measure of noisiness, and not the volume itself, because noisy environments tend to produce volume recordings that are consistently chaotic (ie, high SD), yet the absolute volume of what can be considered *noisy* is often dependent on the microphone and further confounded by device-specific processing, including automatic gain control. This feature was first presented and analyzed in a previous study by Di Matteo et al [30].

Death-Related Words

In addition to the binary presence or absence of speech, the contents of that speech can be investigated to yield further features that may be related to the symptoms of anxiety and depression. The ASR-detected words produced from subjects' audio recordings were analyzed using the Linguistic Inquiry and Word Count (LIWC) software tool [34] to generate a feature that captures relevant linguistic information. LIWC is a tool that can be used to characterize a text along numerous psychological dimensions or word categories. These categories are clusters of words that share a semantic meaning. For example, 1 category that can be explored with LIWC is *affect*, which comprises words relating to emotional states (some examples of the words belonging to the affect category are *happy* and *cried*). For each word category that is of interest, LIWC determines the percentage of words found in the text being analyzed which fall within the predefined dictionary of words for that category. We use LIWC to count the percentage of all words collected from the person that fall within LIWC's *death* category and refer to this as the *death-related words* feature. In our previous work, we showed that the incidence of death-related words has been shown to have an association with symptoms of SAD, GAD, and depression [35]. The previous work of Di Matteo et al [35] also contains further methodological and implementation details of this LIWC-based feature extraction.

Number of Locations Visited

The number of locations that a person visits may also offer insight into mental state, as the avoidance behavior associated with SAD or the lack of energy associated with depression may result in an individual leaving the home less often than healthy individuals. To estimate the number of locations visited from the GPS location data, location data were first processed to identify the stationary points. A stationary point is defined as one in which a subject travels at a speed less than 1 km/h relative to the last recorded location point in time (assuming direct-line travel). This stationary location data were then processed using a clustering approach [36] to group closely located readings into distinct locations visited by the subject. The specific clustering algorithm used is called density-based spatial clustering of applications with noise [37], parameterized with an epsilon value of 150 m. The *number of locations visited* feature is simply the number of clusters produced when clustering a subject's stationary GPS location data. This feature was adopted from the study by Saeb et al [7].

Number of Exits From the Home

The number of times a person leaves their home to visit another location is similarly hypothesized to be relevant to symptoms of anxiety and depression. Although subjects do not explicitly provide us with the location of their home, we infer their home location (or cluster) as the cluster in which participants spend the most amount of time between the hours of midnight and 6 AM, a method that has been used in previous studies [7,11]. Once the home cluster is identified, the *number of exits from the home* can be computed as the number of cluster transitions from the home cluster to any other nonhome cluster.

Screen Use

The *screen use* feature is computed as the proportion of time that each subject's smartphone screen is on. It was computed from the screen on/off time series data. This feature was adopted from the study by Saeb et al [7].

Time in Darkness

The *time in darkness* feature is computed as the proportion of time that each subject's light sensor readings measure an illuminance less than 5 lux, which corresponds to a very dark environment.

Predictive Model Building and Evaluation

Our goal was to assess the capability of this set of 8 features to predict the screening results (ie, above or below a screening threshold) of SAD, GAD, and depression. To do so, logistic regression [38] models were built and independently evaluated for each of the 3 disorders. As a preprocessing step before model training, all 8 features were scaled to have a mean of 0 (SD 1). Any missing feature values (due to data loss) were imputed by median imputation [39]. Feature scaling and imputation were performed in an unbiased fashion, using only *training* data to estimate the mean, SD, and median to perform these operations.

To estimate the predictive performance, a cross-validation [40] scheme was used to provide an unbiased estimate of how well these models predict screening results for unseen subjects (ie, subjects whose data were not used to build or *train* the model). This method is often used in machine learning–based approaches. The cross-validation approach used was repeated *k*-fold cross-validation [41], with 5 folds and 20 repeats. In this approach, the data set is partitioned into 5 *folds* with approximately 17 subjects each. Then, four of those five folds were used to train a logistic regression model, and the accuracy was then evaluated by predicting the unseen screening results of the held-out fold. This process is repeated for a total of 20 times, where in each of the 20 iterations, the data set is split into a different random assignment of participant data to the 5 folds. As every repetition of the 5-fold cross-validation scheme produces 5 models to be trained and evaluated (1 for each of the held-out sets), a total of 100 logistic regression models were trained and evaluated for each of the 3 disorders. A repeated *k*-fold cross-validation scheme was used instead of simple cross-validation because the data set is relatively small and, therefore, the performance of the models can vary by chance depending on how the subjects are randomly assigned to folds.

The predictive performance of each of the 3 groups of models (GAD, SAD, and MDD) is reported as the mean (across the 100 models) of the area under the receiver operating characteristic (AUROC) [42], which is sometimes simply referred to as area under the curve. A modified version of a 1-tailed *t* test was used to assess whether the mean predictive performance of each group of models is significantly better than that of an uninformative model (which would produce random predictions). This corrected repeated *k*-fold cross-validation *t* test was developed by Bouckaert and Frank [43] to account for the fact that the performance of each model produced by a repeated *k*-fold cross-validation resampling is not independent, as models built using this strategy share training data.

Finally, to inspect and interpret the models and the relative influence of each feature within the model, a single model was trained using the entire data set for each of the 3 disorders under investigation. The coefficients of these *full* models will then be presented and discussed. These coefficients are x-standardized as all features were scaled to have mean of 0 (SD 1) before model building. This standardization was necessary because we wish to compare the effects of features that do not share a common unit of measurement or scale.

Ethics and Privacy

All recruited subjects retained anonymity and were provided with randomized account log-ins for use in the study app. All data were transmitted to the study servers over encrypted channels and stored in an encrypted form. Audio recordings were processed entirely by automated software with no human intervention and were deleted immediately upon processing. The words contained in the transcripts of audio recordings were stored in a randomized order to prevent the recreation of speech, after which each transcript was destroyed. The study was approved by the University of Toronto Health Sciences Research Ethics Board (protocol #36687).

Results

Subject Demographics

Of the subjects who completed the study, 75% (84/112) yielded sufficient data for analysis.

Missing data were encountered because of the smartphone app; in some cases, sampling data were obtained at rates below the intended rate. For each participant and for every data stream (ie, audio, GPS, and light), the number of data samples collected in that stream was compared with the number of samples that would be collected given perfect periodic sampling at the desired sampling rate. If the number of samples was less than half of what would be expected, then features would not be computed from that data stream, and those associated features would be considered missing. Participants having four or more missing features were then considered to have insufficient data and were excluded from the analysis.

The gender balance of the study sample was 42% (35/84) female and 58% (49/84) male, with an average age of 28.8 (SD 8.6) years. The 84 subjects included in the analysis and the 28 subjects excluded from the analysis did not differ significantly in mean age, gender distribution, or mean score of any of the 4 self-report measures.

Subject Mental Health Screening

Subjects were screened for SAD, GAD, and depression using the LSAS, GAD-7, and PHQ-8 instruments, respectively. The screening criteria and the number and percentage of positive screenings are summarized in Table 1. The prevalence of SAD, GAD, and MDD in this sample was higher than that in the general Canadian population [30]. We hypothesize that this may be, in part, explained by the fact that study subjects who rely upon Prolific or other crowdsourced work may be under precarious employment, which is linked to poor physical and mental health [44].

Table 1. Subject screening results for social anxiety disorder, generalized anxiety disorder, and depression (n=84).

Disorder	Screening criteria	Positive screenings, n (%)
Social anxiety disorder	LSAS[a] score ≥60	32 (38)
Generalized anxiety disorder	GAD-7[b] score ≥10	22 (26)
Major depressive disorder	PHQ-8[c] score ≥10	31 (37)

[a] LSAS: Liebowitz Social Anxiety Scale.

[b] GAD-7: Generalized Anxiety Disorder 7-item.

[c] PHQ-8: Patient Health Questionnaire 8-item.

Features Extracted From Smartphone-Collected Data

Summary statistics for all eight features extracted from the subjects' smartphone-collected data are presented in Table 2. The daily similarity and weeknight sleep disturbance features do not lend themselves to intuitive interpretation, and we suggest referring to the mathematical definitions provided in the *Methods* section. The remainder of the features are quite simple and can be easily interpreted in the context of everyday behaviors. The speech presence feature has a mean of 0.15, which indicates that, on average, subjects spent 15% of their time (including sleeping time) in the presence of smartphone-detected speech. The mean value of the death-related words feature was 0.16%, indicating that these words were detected infrequently. The mean of the number of locations visited feature was 15, and the mean of the number of exits from the home feature was 15. The subjects' devices had their screens in an on state 23% of the time, on average, and devices were in darkly lit environments 63% of the time.

Table 2. Summary statistics for all subjects' features (n=84).

Feature	Value, mean (SD)	Value, minimum	Value, first quartile	Value, second quartile	Value, third quartile	Value, maximum
Daily similarity	0.80 (0.07)	0.45	0.77	0.83	0.85	0.90
Speech presence	0.15 (0.06)	0.01	0.11	0.15	0.20	0.30
Weeknight sleep disturbance	0.14 (0.06)	0.03	0.09	0.12	0.18	0.34
Death-related words	0.16 (0.10)	0.00	0.09	0.15	0.20	0.51
Locations visited	15 (9.90)	1	8	13	20.5	50
Exits from home	15 (6.90)	1	11	14	18.5	34
Screen use	0.23 (0.13)	0.02	0.12	0.21	0.29	0.60
Time in darkness	0.63 (0.14)	0.31	0.60	0.63	0.74	0.96

Predictive Performance for Models of SAD, GAD, and MDD

Three independent logistic regression models were built and evaluated using all eight features to predict SAD, GAD, and depression. As noted in the *Methods* section, a repeated *k*-fold cross-validation scheme was used to acquire an unbiased estimate of the models' predictive performance. The mean AUROC for each of the 3 disorder models is shown in Table 3. The SAD model achieved a mean AUROC of 0.64 (SD 0.13), the GAD model achieved a mean AUROC of 0.55 (SD 0.14), and the depression model achieved the best results with a mean AUROC of 0.72 (SD 0.12).

Table 3. Aggregate predictive performance of the resampled models of social anxiety disorder, generalized anxiety disorder, and depression.

Disorder model	AUROC[a], mean (SD)	*t* test (*df*)	*P* value
SAD[b]	0.64 (0.13)	2.04 (99)	.02
GAD[c]	0.55 (0.14)	0.73 (99)	.23
MDD[d]	0.72 (0.12)	3.46 (99)	<.001

[a] AUROC: area under the receiver operating characteristic.

[b] SAD: social anxiety disorder.

[c] GAD: generalized anxiety disorder.

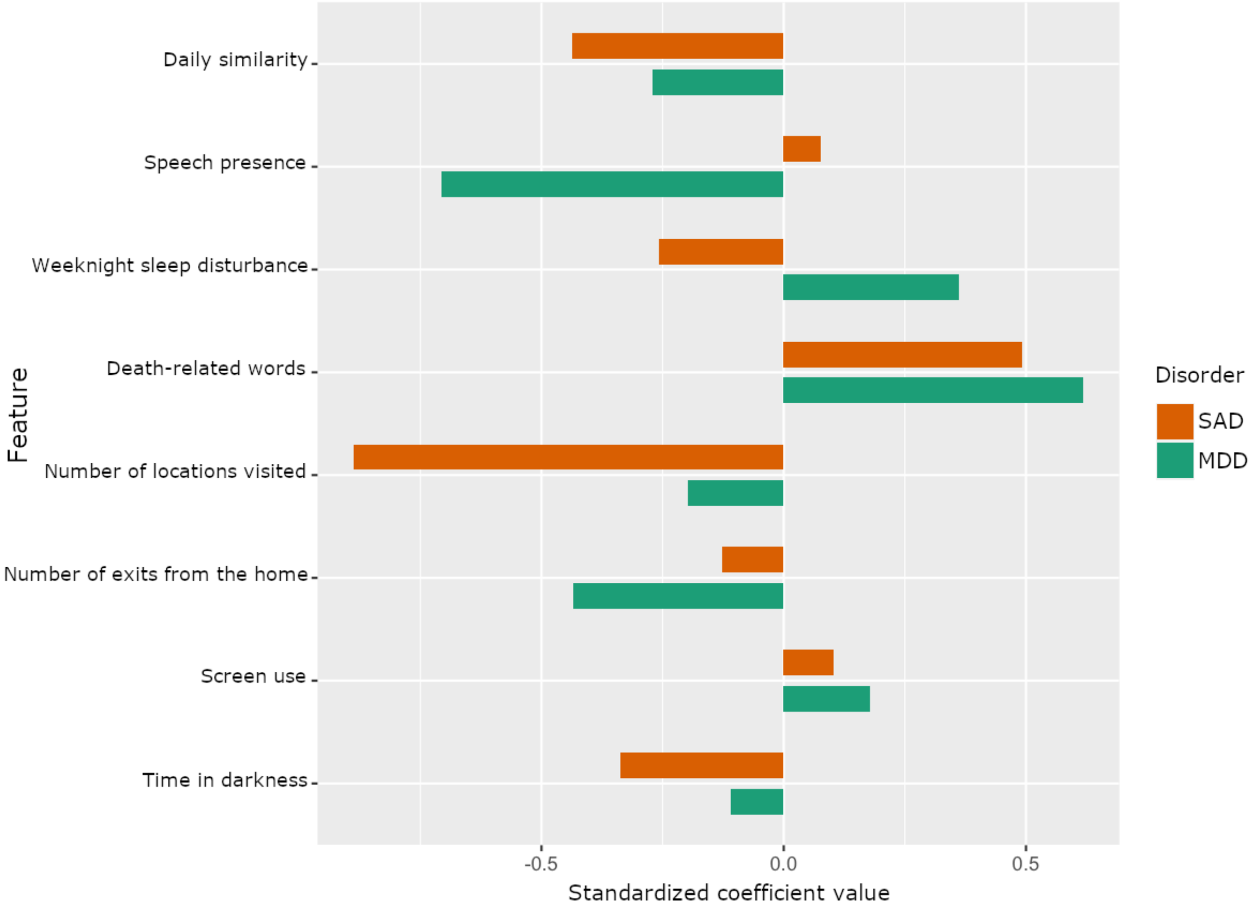
[d] MDD: major depressive disorder.

Table 3 also provides the results of a modified *t* test to compare whether the mean AUROC of each model was significantly better than that of an uninformative model (which performs with AUROC=0.5). Both the SAD model and the depression model performed significantly better than the uninformative model at a 5% significance level. The GAD model effectively performed on par with an uninformative model (mean AUROC 0.55), showing no effective capability to predict GAD.

Feature Importance

We sought to determine the relative importance of the features and to compare how feature importance may differ between the models for the 3 different disorders. The standardized logistic regression coefficients of the models of SAD and depression are shown in Figure 2. The GAD model was not included because it failed to achieve a significant predictive capability. As described in the *Methods* section, these coefficients come from a *single* model that was built on the entire data set of 84 subjects for each disorder, in contrast with the results in Table 3, for which many models' performances were averaged (in the cross-validation scheme). The difference here is that the aim is not to measure performance but to inspect the model characteristics; hence, only 1 model per disorder, trained on the entire data set, is needed.

Figure 2. Comparison of logistic regression model coefficients by disorder. MDD: major depressive disorder; SAD: social anxiety disorder.



First, by looking at the signs of the feature coefficients, we can determine which features were found to decrease the odds of a positive screening, as negative coefficients correspond to a feature that is associated with *decreased* odds of a positive screening (a characteristic of the logistic regression modeling used). The following features were associated with decreased odds of a positive screen for both social anxiety and depression: daily similarity, number of locations visited, number of exits from the home, and time in darkness. This indicates that these features generally capture healthy behaviors with respect to social anxiety and depression. In contrast, the death-related words and screen use features were both associated with increased odds of positive screening for SAD and depression.

Two features, namely, speech presence and weeknight sleep disturbance, were found to have contrasting directionality with respect to screening for SAD and depression. Speech presence was associated with greatly *decreased* odds of a depression screening; however, speech presence was associated with a slight increase in the odds of a SAD screening. In addition, weeknight sleep disturbance was associated with increased odds of a depression screening but with decreased odds of a SAD screening.

Second, in terms of the relative magnitudes of their associated coefficients, several features stand out. For SAD, the daily similarity and number of locations visited features seem to be the strongest indicators of health, whereas the death-related words feature is the greatest risk factor for SAD. For depression, the speech presence and number of exits from the home features seem to be the strongest indicators of health, whereas weeknight sleep disturbance and death-related words features are the greatest risk factors for depression.

Discussion

Modeling and Feature Importance

The moderate success of the predictive models of SAD and depression is interesting when compared with the failure of the GAD model. This particular feature set seems unable to predict GAD, which may be related to how generalized anxiety manifests itself in subjects' feelings and behaviors in a manner different from social anxiety or depression. Individuals with SAD commonly use behavioral avoidance to avoid situations that trigger their anxiety symptoms; they may choose specific jobs to avoid public speaking or avoid going to social gatherings [45]. The relative lack of or decrease in public speaking, traveling, and leaving the home are all physical behaviors that can be detected by some of the features used in this study. The speech presence feature can measure changes in the amount of speech, the number of locations visited feature can measure changes in the amount of travel, and the number of exits from the home feature directly measures how often one leaves the home. Although individuals with depression do not avoid specific situations for the same reasons that individuals with social anxiety do, depression is characterized by a general lack of motivation and energy [46]. This lack of motivation and energy also appears to manifest itself in behaviors (or lack thereof), which are detectible by our feature set (specifically the speech presence, number of locations visited, and number of exits from the home features).

This *behavioral* avoidance can be contrasted with the *cognitive* avoidance that is common in individuals with GAD [47], which includes maladaptive and somewhat pathological strategies such as distraction, worry, and thought suppression [48]. These are all strategies used by individuals with GAD to suppress their symptoms; however, they do not as easily manifest in the physical realm in a way that can be detected through, for example, GPS location data. In other words, it may be that this study's feature set does a good job of detecting behaviors but a poor job in inferring the cognitive strategies that may often be used in a pathological manner to combat negative and, therefore, aversive feelings. Although it is true that screen use–based features could infer times when subjects were using distraction as a behavioral avoidance strategy, there are other possible explanations for screen time that might not include avoidance. The screen use data that we collected did not contain information on what apps were being used while the screen was on, which could be used to determine the motivation for using the phone. This more fine-grained use data that detailed which apps were in use (eg, games vs productivity apps such as email) would yield more insight into this type of avoidance and therefore may be more predictive of GAD.

A similar line of reasoning was presented in an earlier study of the audio-based features and how they were less strongly correlated with GAD symptom severity than the symptom severity of SAD and MDD [30]. It is worth noting that the additional features presented in this work (which were derived from GPS location, screen data, and light sensor data) also fail to effectively predict GAD.

Comparing the Models of SAD and MDD

The predictive models of SAD and depression have similar performance, achieving mean AUROC values of 0.64 (SD 0.13) and 0.72 (SD 0.12), respectively. Numerous features appear to be of near-equal importance for both models, whereas some features are much more important for one of the two disorders (SAD and MDD). Comparing the 2 models, the proportion of death-related words detected in the environmental audio recordings was the greatest risk factor for both SAD and depression. The link between the use of death-related words and depression has been demonstrated in some previous studies [49,50], but we are not aware of any empirical studies that have demonstrated a link between death-related words and symptoms of SAD. Some researchers have proposed that death anxiety and fear of death may function as a transdiagnostic construct underpinning a range of mental disorders, including anxiety disorders [51]. If this hypothesis holds true, then the presence of death-related words in environmental audio may also serve as a proxy measure of death anxiety.

Continuing to compare the models of SAD and MDD, the daily similarity feature appears as a factor of good health with respect to both SAD and depression, in line with the hypotheses outlined in the description of this feature in the *Methods* section. That is, higher regularity in the subjects' patterns of daily activities may be associated with better mental health.

A difference between the 2 models is that 2 features appear to act as predictors in opposite directions. First, the speech presence feature appears to be a critical indicator of health with respect to depression: subjects who spend more time in the presence of speech in their environment had greatly reduced odds of depression. The same is not true for SAD, in which the direction of this effect was reversed. Nonetheless, the value of the coefficient for speech presence in the SAD model is so low that it is considered an insignificant risk factor. The result that links more environmental speech to weaker symptoms of depression replicates results from previous studies [12,15]. It is interesting to note that more environmental speech was not associated with weaker symptoms of SAD. It may be the case that more information regarding whom the subject is speaking to (assuming that the speech is from present humans and not prerecorded speech from a device of some kind), and in what setting, is necessary to identify the key contexts in which social interactions are relevant to a subject's symptoms of SAD.

The weeknight sleep disturbance feature is the second of 2 features that act in different directions within the models of SAD and depression. Sleep disturbance is a risk factor in the model of depression, a result that aligns with our original hypothesis. However, sleep disturbance was shown to decrease the odds of a positive screening of SAD. This indicates that sleep disturbance may be a positive factor with respect to social anxiety, if enhanced nighttime activity is indeed the result of social interaction.

Finally, there are 2 features that serve as positive factors for both SAD and depression, but with marked differences in magnitude. The number of locations visited feature is much more impactful in the model of SAD than in the model of depression, whereas the number of exits from the home feature is much more impactful for depression than for SAD. This may suggest that for depression, simply leaving the home and engaging in some form of activity is enough to preclude a positive screening. For SAD, simply leaving the home may not provide enough opportunity for social engagement to indicate a lack of social anxiety; however, visits to numerous locations (some of which may be social in nature) may suggest that an individual does not have SAD.

Comparison With Other Studies

A number of studies have used smartphone-collected data in a similar fashion to build and evaluate predictive models of depression. Saeb et al [7] achieved a classification accuracy of 78.8% in detecting individuals with depression (which was defined as having a PHQ-9 score ≥5) using only location-based features. To enable more direct comparisons, 2 existing studies have also reported AUROC as their metric of accuracy in predicting depression using smartphone-collected data. Wang et al [52] reported an AUROC of 0.81 for a model that predicted depression using location, audio, and screen-based features. A study by Place et al [53] achieved an AUROC of 0.74 for detecting depressed mood using features derived from GPS location, audio, motion sensor data, phone and messaging metadata, screen data, and other device data.

Both the studies by Wang et al [52] and Place et al [53] differ from ours, however, in how they screen subjects for depression, as they used scores from the abbreviated 2-item PHQ-2 depression instrument. Furthermore, in both studies, participants could be considered to be drawn from a more homogenous sample, as they were both geolocated in particular metro areas. The participants in the study by Wang et al [52] were 48 undergraduate students at Dartmouth College, whereas the 73 participants in the study by Place et al [53] were all residents of the Boston area. The participants in our study were a mix of students and nonstudents and were geographically located across Canada in both rural and urban areas. The nature of the area in which subjects live and travel is of particular importance for location-based features, as individuals living in rural areas may exhibit different patterns of travel than those in urban areas.

Studies of anxiety disorders using smartphone-collected data are much less represented in the literature. A study by Boukechba et al [10] demonstrated strong correlations between smartphone-collected data and symptom severity of SAD, but no classification (ie, predictive screening) was performed. To our knowledge, there are currently no studies that have predicted or otherwise measured correlations between GAD and smartphone-collected data.

Limitations

One limitation of this study is that the subjects' state of SAD, GAD, and depression was determined using self-report measures as screeners. Although all the instruments used demonstrated acceptable accuracy for doing so, it is not clear how these results would compare had the diagnoses been performed using structured clinical interviews by trained clinicians.

Another limitation is related to the speech presence feature, which captures any intelligible speech within range of participants' devices, including television or other media. More information about whom the subject is speaking with (ie, humans present in the environment, in comparison with recorded speech, be it from television, radio, etc) and where the individual is situated at the time of the recording may improve the ability to identify the key contexts in which social interactions are relevant to a subject's symptoms of SAD.

Another set of key limitations involves the use of a particular smartphone system to collect data. First, all subjects were users of Android smartphones. Although our study excluded iOS users, 1 study of depression prediction from smartphone data that separately considered both iOS and Android users found similar results for both subgroups [14]. Second, the features derived from the smartphone-collected data in this study have some limitations. The speech presence feature and the death-related words features do not distinguish between speech produced by the subject and speech produced by other people. The screen use feature does not account for which app, if any, the subject is interacting with while the screen is on. Finally, the time in darkness feature does not distinguish between the device measuring low light conditions because the subject is in a dark room (eg, while sleeping) or the device simply being in a dark location (eg, in a pocket).

Conclusions

This work contributes to the development of automated technology capable of screening individuals for SAD, GAD, and depression. To our knowledge, this is the first study to evaluate and compare how the same set of predictor variables derived from objective smartphone-collected data perform in predicting depression and 2 classes of anxiety disorders. This set of features was used to build models that had a significant capability in predicting SAD (mean AUROC 0.64, SD 0.13; P=.02) and depression (mean AUROC 0.72, SD 0.12; P<.001). Although the prediction of GAD was unsuccessful, we believe that we are the first to evaluate that disorder using this methodology and to propose key considerations that may yield success in predicting that disorder from smartphone-collected data.

Acknowledgments

This research was funded in part by the National Sciences and Engineering Research Council of Canada Discovery Grant (RGPIN-2019-04395).

Conflicts of Interest

MAK has been a consultant or advisory board member for GlaxoSmithKline, Lundbeck, Eli Lilly, Boehringer Ingelheim, Organon, AstraZeneca, Janssen, Janssen-Ortho, Solvay, Bristol-Myers Squibb, Shire, Sunovion, Pfizer, Purdue, Merck, Astellas, Tilray, Bedrocan, Takeda, Eisai, and Otsuka. MAK has undertaken research for GlaxoSmithKline, Lundbeck, Eli Lilly, Organon, AstraZeneca, Jannsen-Ortho, Solvay, Genuine Health, Shire, Bristol-Myers Squibb, Takeda, Pfizer, Hoffman La Rosche, Biotics, Purdue, Astellas, Forest, and Lundbeck. MAK has received honoraria from GlaxoSmithKline, Lundbeck, Eli Lilly, Boehringer Ingelheim, Organon, AstraZeneca, Janssen, Janssen-Ortho, Solvay, Bristol-Myers Squibb, Shire, Sunovion, Pfizer, Purdue, Merck, Astellas, Bedrocan, Tilray, Allergan, and Otsuka. MAK has received research grants from the Canadian Institutes of Health Research, Sick Kids Foundation, Centre for Addiction and Mental Health Foundation, Canadian Psychiatric Research Foundation, Canadian Foundation for Innovation, and the Lotte and John Hecht Memorial Foundation.

References

1. Smetanin P, Stiff D, Briante C, Adair CE, Ahmad S, Khan M. The life and economic impact of major mental illnesses in Canada: 2011 to 2041. RiskAnalytica, on behalf of the Mental Health Commission of Canada. 2011. URL: https://www.mentalhealthcommission.ca/sites/default/files/MHCC_Report_Base_Case_FINAL_ENG_0_0.pdf [accessed 2020-11-22]
2. Mental health care needs, 2018. Government of Canada. 2019 Oct 07. URL: https://www150.statcan.gc.ca/n1/pub/82-625-x/2019001/article/00011-eng.htm [accessed 2020-11-22]
3. Graham S, Depp C, Lee EE, Nebeker C, Tu X, Kim H, et al. Artificial intelligence for mental health and mental illnesses: an overview. Curr Psychiatry Rep 2019 Nov 07;21(11):116 [FREE Full text] [doi: 10.1007/s11920-019-1094-0] [Medline: 31701320]
4. Benca RM, Okawa M, Uchiyama M, Ozaki S, Nakajima T, Shibui K, et al. Sleep and mood disorders. Sleep Med Rev 1997 Nov;1(1):45-56. [doi: 10.1016/s1087-0792(97)90005-8]
5. Bueno-Antequera J, Munguía-Izquierdo D. Exercise and depressive disorder. Adv Exp Med Biol 2020;1228:271-287. [doi: 10.1007/978-981-15-1792-1_18] [Medline: 32342464]
6. Chen Z, Lin M, Chen F, Lane N, Cardone G, Wang R, et al. Unobtrusive sleep monitoring using smartphones. In: Proceedings of the 7th International Conference on Pervasive Computing Technologies for Healthcare. 2013 Presented at: 7th International Conference on Pervasive Computing Technologies for Healthcare; 2013; Venice, Italy p. 13-152. [doi: 10.4108/icst.pervasivehealth.2013.252148]
7. Saeb S, Zhang M, Karr CJ, Schueller SM, Corden ME, Kording KP, et al. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. J Med Internet Res 2015;17(7):e175 [FREE Full text] [doi: 10.2196/jmir.4273] [Medline: 26180009]
8. Demasi O, Aguilera A, Recht B. Detecting change in depressive symptoms from daily wellbeing questions, personality, and activity. In: Proeedings of the IEEE Wireless Health (WH) Conference. 2016 Presented at: IEEE Wireless Health (WH) Conference; Oct. 25-27, 2016; Bethesda, Maryland, USA. [doi: 10.1109/WH.2016.7764552]
9. Leichsenring F, Leweke F. Social anxiety disorder. N Engl J Med 2017 Jun 08;376(23):2255-2264. [doi: 10.1056/NEJMcp1614701] [Medline: 28591542]
10. Boukhechba M, Chow P, Fua K, Teachman BA, Barnes LE. Predicting social anxiety from global positioning system traces of college students: feasibility study. JMIR Ment Health 2018 Jul 04;5(3):e10101 [FREE Full text] [doi: 10.2196/10101] [Medline: 29973337]
11. Chow PI, Fua K, Huang Y, Bonelli W, Xiong H, Barnes LE, et al. Using mobile sensing to test clinical models of depression, social anxiety, state affect, and social isolation among college students. J Med Internet Res 2017 Mar 03;19(3):e62 [FREE Full text] [doi: 10.2196/jmir.6820] [Medline: 28258049]
12. 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]
13. Canzian L, Musolesi M. Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2015 Presented at: UbiComp '15: The 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 7 - 11, 2015; Osaka, Japan p. 1293-1304. [doi: 10.1145/2750858.2805845]
14. Farhan A, Yue C, Morillo R, Ware S, Lu J, Bi J, et al. Behavior vs. introspection: refining prediction of clinical depression via smartphone sensing data. In: Proceedings of the IEEE Wireless Health (WH) Conference. 2016 Presented at: IEEE Wireless Health (WH) Conference; Oct. 25-27, 2016; Bethesda, MD, USA. [doi: 10.1109/wh.2016.7764553]
15. 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: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2014 Presented at: UbiComp '14: The 2014 ACM Conference on Ubiquitous Computing; September 13 - 17, 2014; Seattle, Washington, USA p. 3-14. [doi: 10.1145/2632048.2632054]

16. Rohani DA, Faurholt-Jepsen M, Kessing LV, Bardram JE. Correlations between objective behavioral features collected from mobile and wearable devices and depressive mood symptoms in patients with affective disorders: systematic review. JMIR Mhealth Uhealth 2018 Aug 13;6(8):e165 [FREE Full text] [doi: 10.2196/mhealth.9691] [Medline: 30104184]
17. Tiller JW. Depression and anxiety. Med J Aust 2013 Sep 16;199(S6):28-31. [doi: 10.5694/mja12.10628] [Medline: 25370281]
18. Online participant recruitment for surveys and market research. Prolific. URL: https://www.prolific.co/ [accessed 2020-01-21]
19. Marcano BJ, 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]
20. Liebowitz MR. Social phobia. Mod Probl Pharmacopsychiatry 1987 Jan 01;22:141-173. [doi: 10.1159/000414022] [Medline: 2885745]
21. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
22. Kroenke K, Strine TW, Spitzer RL, Williams JB, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord 2009 Apr;114(1-3):163-173. [doi: 10.1016/j.jad.2008.06.026] [Medline: 18752852]
23. Leon AC, Olfson M, Portera L, Farber L, Sheehan DV. Assessing psychiatric impairment in primary care with the Sheehan Disability Scale. Int J Psychiatry Med 1997;27(2):93-105. [doi: 10.2190/T8EM-C8YH-373N-1UWD] [Medline: 9565717]
24. Mennin DS, Fresco DM, Heimberg RG, Schneier FR, Davies SO, Liebowitz MR. Screening for social anxiety disorder in the clinical setting: using the Liebowitz Social Anxiety Scale. J Anx Disord 2002 Jan;16(6):661-673. [doi: 10.1016/S0887-6185(02)00134-2]
25. Rytwinski NK, Fresco DM, Heimberg RG, Coles ME, Liebowitz MR, Cissell S, et al. Screening for social anxiety disorder with the self-report version of the Liebowitz Social Anxiety Scale. Depress Anxiety 2009 Jan;26(1):34-38. [doi: 10.1002/da.20503] [Medline: 18781659]
26. Speech-to-Text: Automatic speech recognition. Google Cloud. URL: https://cloud.google.com/speech-to-text/ [accessed 2020-01-12]
27. Optimize for doze and app standby. Android Developers. URL: https://developer.android.com/training/monitoring-device-state/doze-standby [accessed 2020-06-08]
28. Shear MK, Randall J, Monk TH, Ritenour A, Tu X, Frank E, et al. Social rhythm in anxiety disorder patients. Anxiety 1994 Nov 07;1(2):90-95. [doi: 10.1002/anxi.3070010208] [Medline: 9160553]
29. Szuba MP, Yager A, Guze BH, Allen EM, Baxter LR. Disruption of social circadian rhythms in major depression: a preliminary report. Psychiatry Res 1992 Jun;42(3):221-230. [doi: 10.1016/0165-1781(92)90114-i]
30. Di Matteo D, Fotinos K, Lokuge S, Yu J, Sternat T, Katzman MA, et al. The relationship between smartphone-recorded environmental audio and symptomatology of anxiety and depression: exploratory study. JMIR Form Res 2020 Aug 13;4(8):e18751 [FREE Full text] [doi: 10.2196/18751] [Medline: 32788153]
31. Craske M. Chapter 5 - Anxious processes. In: Craske M, editor. Origins of Phobias and Anxiety Disorders. Amsterdam: Elsevier Science; Nov 13, 2003:71-98.
32. Trew JL. Exploring the roles of approach and avoidance in depression: an integrative model. Clin Psychol Rev 2011 Nov;31(7):1156-1168. [doi: 10.1016/j.cpr.2011.07.007] [Medline: 21855826]
33. Neckelmann D, Mykletun A, Dahl AA. Chronic insomnia as a risk factor for developing anxiety and depression. Sleep 2007 Jul;30(7):873-880 [FREE Full text] [doi: 10.1093/sleep/30.7.873] [Medline: 17682658]
34. Tausczik YR, Pennebaker JW. The psychological meaning of words: LIWC and computerized text analysis methods. J Lang Soc Psychol 2009 Dec 08;29(1):24-54. [doi: 10.1177/0261927X09351676]
35. Di Matteo D, Wang W, Fotinos K, Lokuge S, Yu J, Sternat T, et al. Smartphone-detected ambient speech and self-reported measures of anxiety and depression: exploratory observational study. JMIR Form Res 2021 Jan 29;5(1):e22723 [FREE Full text] [doi: 10.2196/22723] [Medline: 33512325]
36. Badillo S, Banfai B, Birzele F, Davydov II, Hutchinson L, Kam-Thong T, et al. An introduction to machine learning. Clin Pharmacol Ther 2020 Apr;107(4):871-885 [FREE Full text] [doi: 10.1002/cpt.1796] [Medline: 32128792]
37. Ester M, Kriegel H, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. 1996 Aug 02 Presented at: Second International Conference on Knowledge Discovery and Data Mining (KDD); August 2 - 4, 1996; Portland, Oregon, USA p. 226-231. [doi: 10.5555/3001460.3001507]
38. Stoltzfus JC. Logistic regression: a brief primer. Acad Emerg Med 2011 Oct;18(10):1099-1104 [FREE Full text] [doi: 10.1111/j.1553-2712.2011.01185.x] [Medline: 21996075]
39. Zhang Z. Missing data imputation: focusing on single imputation. Ann Transl Med 2016 Jan;4(1):9 [FREE Full text] [doi: 10.3978/j.issn.2305-5839.2015.12.38] [Medline: 26855945]
40. Bzdok D, Meyer-Lindenberg A. Machine learning for precision psychiatry: opportunities and challenges. Biol Psychiatry Cogn Neurosci Neuroimaging 2018 Dec;3(3):223-230. [doi: 10.1016/j.bpsc.2017.11.007] [Medline: 29486863]
41. Bischl B, Mersmann O, Trautmann H, Weihs C. Resampling methods for meta-model validation with recommendations for evolutionary computation. Evolut Comput 2012 Jun;20(2):249-275. [doi: 10.1162/evco_a_00069]

42. Obuchowski NA, Bullen JA. Receiver operating characteristic (ROC) curves: review of methods with applications in diagnostic medicine. Phys Med Biol 2018 Mar 29;63(7):07TR01. [doi: 10.1088/1361-6560/aab4b1] [Medline: 29512515]
43. Bouckaert R, Frank E. Evaluating the replicability of significance tests for comparing learning algorithms. In: Dai H, Srikant R, Zhang C, editors. Advances in Knowledge Discovery and Data Mining. Berlin: Springer; 2004:3-12.
44. Benach J, Vives A, Amable M, Vanroelen C, Tarafa G, Muntaner C. Precarious employment: understanding an emerging social determinant of health. Annu Rev Public Health 2014;35:229-253. [doi: 10.1146/annurev-publhealth-032013-182500] [Medline: 24641559]
45. Beidel D, Turner S. Shy Children, Phobic Adults: NatureTreatment of Social Anxiety Disorders, 2nd Ed. Washington, DC, USA: American Psychological Association; 2006:1-398.
46. Cooper JA, Arulpragasam AR, Treadway MT. Anhedonia in depression: biological mechanisms and computational models. Curr Opin Behav Sci 2018 Aug;22:128-135 [FREE Full text] [doi: 10.1016/j.cobeha.2018.01.024] [Medline: 29503842]
47. Borkovec TD, Alcaine OM, Behar E. Avoidance theory of worry and generalized anxiety disorder. In: Heimberg R, Turk C, Mennin D, Walters E, Wittchen H, editors. Generalized Anxiety Disorder: Advances in Research and Practice. New York City, USA: The Guilford Press; 2004:1-446.
48. Mahoney AE, Hobbs MJ, Newby JM, Williams AD, Sunderland M, Andrews G. The worry behaviors inventory: assessing the behavioral avoidance associated with generalized anxiety disorde. J Affect Disord 2016 Oct;203:256-264. [doi: 10.1016/j.jad.2016.06.020] [Medline: 27314812]
49. Veltman B. Linguistic analysis of the semantic content of the Rorschach Inkblot Test. ProQuest. 2006. URL: https://www.proquest.com/docview/304953299 [accessed 2020-12-24]
50. 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]
51. Iverach L, Menzies RG, Menzies RE. Death anxiety and its role in psychopathology: reviewing the status of a transdiagnostic construct. Clin Psychol Rev 2014 Nov;34(7):580-593 [FREE Full text] [doi: 10.1016/j.cpr.2014.09.002] [Medline: 25306232]
52. 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]
53. Place S, Blanch-Hartigan D, Rubin C, Gorrostieta C, Mead C, Kane J, et al. Behavioral indicators on a mobile sensing platform predict clinically validated psychiatric symptoms of mood and anxiety disorders. J Med Internet Res 2017 Mar 16;19(3):e75 [FREE Full text] [doi: 10.2196/jmir.6678] [Medline: 28302595]

Abbreviations

ASR: automatic speech recognition
AUROC: area under the receiver operating characteristic
GAD-7: Generalized Anxiety Disorder 7-item
GAD: generalized anxiety disorder
LIWC: Linguistic Inquiry and Word Count
LSAS: Liebowitz Social Anxiety Scale
MDD: major depressive disorder
PHQ-8: Patient Health Questionnaire 8-item
SAD: social anxiety disorder

Edited by R Kukafka; submitted 18.03.21; peer-reviewed by B Teachman, SC Guntuku, J Kane; comments to author 14.05.21; revised version received 30.06.21; accepted 05.07.21; published 13.08.21.

Please cite as:
Di Matteo D, Fotinos K, Lokuge S, Mason G, Sternat T, Katzman MA, Rose J
Automated Screening for Social Anxiety, Generalized Anxiety, and Depression From Objective Smartphone-Collected Data: Cross-sectional Study
J Med Internet Res 2021;23(8):e28918
URL: https://www.jmir.org/2021/8/e28918
doi:10.2196/28918
PMID:34397386

©Daniel Di Matteo, Kathryn Fotinos, Sachinthya Lokuge, Geneva Mason, Tia Sternat, Martin A Katzman, Jonathan Rose. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Next-Generation Digital Biomarkers for Tuberculosis and Antibiotic Stewardship: Perspective on Novel Molecular Digital Biomarkers in Sweat, Saliva, and Exhaled Breath

Noe Brasier[1,2], MD; Michael Osthoff[3], MD; Fiorangelo De Ieso[1,3], MSc; Jens Eckstein[1,3], MD, PhD

[1]Department of Digitalization & ICT, University Hospital Basel, Basel, Switzerland

[2]Institute for Translational Medicine, ETH Zurich, Zurich, Switzerland

[3]Division of Internal Medicine, University Hospital Basel, Basel, Switzerland

Corresponding Author:
Noe Brasier, MD
Department of Digitalization & ICT
University Hospital Basel
Markgräflerhof / Innovation Lab Hebelstrasse 10
Basel, 4056
Switzerland
Phone: 41 797025837
Email: noe.brasier@usb.ch

Abstract

The internet of health care things enables a remote connection between health care professionals and patients wearing smart biosensors. Wearable smart devices are potentially affordable, sensitive, specific, user-friendly, rapid, robust, lab-independent, and deliverable to the end user for point-of-care testing. The datasets derived from these devices are known as digital biomarkers. They represent a novel patient-centered approach to collecting longitudinal, context-derived health insights. Adding automated, analytical smartphone applications will enable their use in high-, middle-, and low-income countries. So far, digital biomarkers have been focused primarily on accelerometer data and heart rate due to well-established sensors originating from the consumer market. Novel emerging smart biosensors will detect biomarkers (or compounds) independent of a lab and noninvasively in sweat, saliva, and exhaled breath. These molecular digital biomarkers are a promising novel approach to reduce the burden from 2 major infectious diseases with urgent unmet needs: tuberculosis and infections with multidrug resistant pathogens. Active tuberculosis (aTbc) is one of the deadliest diseases from an infectious agent. However, a simple and reliable test for its detection is still missing. Furthermore, inappropriate antimicrobial use leads to the development of antimicrobial resistance, which is associated with high mortality and health care costs. From this perspective, we discuss the innovative approach of a noninvasive and lab-independent collection of novel biomarkers to detect aTbc, which at the same time may additionally serve as a scalable therapeutic drug monitoring approach for antibiotics. These molecular digital biomarkers are next-generation digital biomarkers and have the potential to shape the future of infectious diseases.

(J Med Internet Res 2021;23(8):e25907) doi:10.2196/25907

KEYWORDS

digital biomarkers; active tuberculosis; drug resistance; wearable; smart biosensors; iSudorology; infectious diseases

Introduction

A biomarker is defined as "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention" by the National Institutes of Health Biomarkers Definitions Working Group [1]. Biomarkers are mostly collected from standard biofluids such as blood and urine, among others [2,3]. Compared with unspecific inflammatory markers such as C-reactive protein (CRP), more specific biomarkers such as bacterial culture support diagnosis and disease monitoring [2]. The emergence of smart devices as part of the internet of health care things allows for connecting a patient with health care workers in a location- and lab-independent way [4]. Recording a patient's health status independent of a lab and remotely provides a novel entity for health and disease data, referred to as a digital biomarker [5]. Smart biosensors collect digital biomarkers such as accelerometer data, heart rate, and body temperature mostly in a noninvasive way. By connecting smart biosensors to automated analytical smartphone applications, digital biomarkers are disrupting standard operating procedures in patient monitoring [6]. The data collected are subsequently stored, shareable, and therefore serve as an outbreak detection network for infectious diseases as described by Tom-Aba et al [7].

Smart biosensors from the consumer market that monitor a patient's activity, electrocardiogram, heart rate, and other parameters are increasingly available as certified medical devices [8]. The potential to continuously record patient-centered information from daily life will provide deeper insight into health and disease. Investigations into digital biomarkers have been focused on health measurements such as activity and heart rate due to the broad availability of accelerometer sensors. Patients with multiple sclerosis as well as those with Parkinson's disease–related tremors are subjects of current clinical research [9,10], and studies are also investigating smart device–based automated algorithms to screen for atrial fibrillation using photoplethysmographic rhythm analysis [11].

Point-of-care testing (POCT) provides personalized and actionable case identification near a patient's location [12]. POCT needs to match the ASSURED criteria according to the World Health Organization (WHO): (1) affordable, (2) sensitive, (3) specific, (4) user-friendly, (5) rapid and robust, (6) equipment-free, and (7) deliverable to the end user [13]. POCT supports health care professionals in high-income countries, as well in middle- and low-income countries with restricted health care infrastructure [14,15]. The most significant barriers for broad use of POCT are lack of training, increased patient waiting time, and lack of availability [16]. Wearable devices linked to automated smartphone applications can be applied at home irrespective of the distance to the nearest health care facility [17]. Novel biosensors will provide the next generation of digital biomarkers by noninvasively detecting molecular feedback [18,19]. The increasingly available smartphone-based biosensors meet the WHO ASSURED criteria and will shape the future of managing infectious diseases around the globe.

From this perspective, we explored novel smartphone-based biosensors to analyze sweat, saliva, and exhaled breath for 2 of the main global issues in infectious diseases. We discuss the potential of molecular digital biomarkers in patients with active tuberculosis (aTbc), 1 of the 3 important epidemics [20]. In addition, we will give a perspective on the use of smart biosensors for noninvasive, personalized, pharmacologic monitoring in patients treated with antibiotics. Antibiotic drug resistance is a threat to global health and a major public health issue [21]. Finally, we will summarize the remaining limitations that have prevented these approaches from being clinically implemented.

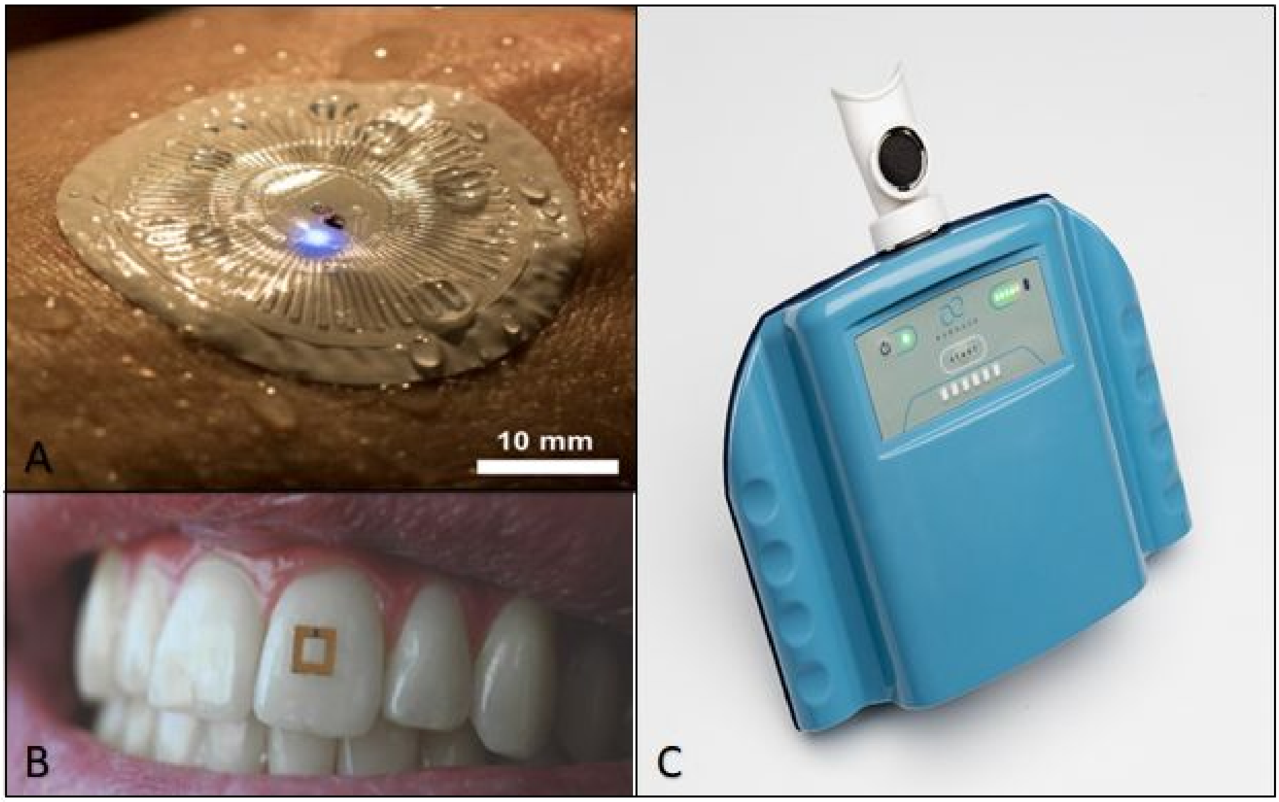
Novel Biosensors for Sweat, Saliva, and Exhaled Breath Analysis

Biofluids such as sweat, saliva, and exhaled breath are noninvasively collectable and represent a promising pool of continuously available molecular biomarkers. So far, sweat, saliva, and exhaled-breath samples have mostly been analyzed using laborious and expensive mass spectrometry [22-24]. Smart biosensors coupled with smartphone applications are increasingly available and allow for lab-independent analysis [25-27]. Smart biosensors connect patients and health care professionals remotely, mostly through Bluetooth or WiFi and may have a global impact in high- as well as in middle- and low-income countries with less available health care infrastructure [4].

Sweat Sensors

Sweat samples are often restricted to low-volume, low-molecule concentrations and need sample stabilization. Noninvasive, lab-independent, on-skin sweat analysis is increasingly available. There are electrochemical-based [28] (Figure 1A), enzyme-linked immunoassay (ELISA)–based [29], and aptameric-based sensors [30], among others. The clinical implementation of sweat biosensors has been restricted, mostly due to unsolved challenges such as sensor stability and low biomarker concentrations. The first clinical tests were successfully conducted using a sweat sensor that detects uric acid as well as tyrosine as described by Yang et al [31].

Figure 1. Next–generation biosensors in health care: (A) on-skin sweat biosensor for sweat analysis by Reeder et al [27]; (B) tooth-mounted biosensor for saliva analysis by Tseng et al [25]; and (C) e-nose for exhaled breath analysis by Aeonose [32]. Permission to reproduce the images was obtained from the respective publishers/authors.



Saliva Sensors

A biosensor for in-mouth analysis of saliva has been developed by Tseng et al [25] (Figure 1B). This small sensor is designed to be tooth-mounted and allows for continuous, lab-independent saliva analysis. The functionality of this trilayer sensor was tested in vivo by detecting glucose in different conditions such as dry mouth, after drinking tap water, and after drinking apple juice. Analysis is conducted in combination with a portable vector analyzer attached to a tablet or cell phone. Reliable and stable detection of glucose in vivo was demonstrated for more than 1 week.

Exhaled Breath Sensors

Biosensors to analyze exhaled breath can be summarized under the term “electric nose” (EN) [26] (Figure 1C). The applicability ranges from health care [33], the food and beverage industry [34], to general monitoring [35]. EN can be compared to the human sense of smell, as the sensor array represents the human nose and data analysis represents the human brain. There are different sensor entities such as low-cost metal oxide sensors performing well at high temperatures; fast and reliable conducting polymer sensors responding to odors; and sensitive, rapid, and stable quartz crystal microbalance [26]. EN has already been clinically validated [36].

Perspective on Smart Biosensors

Smart biosensors are promising novel approaches to overcome the current remote patient monitoring issues such as lab dependence, laborious sample preparation, and time delays. As analysis can be automated, biosensors will be intuitively easy to handle. The novel opportunity to record continuous molecular feedback will provide novel high-resolution information about health and disease. Nevertheless, easy and continuous health measurements lead to novel issues such as massive amounts of recorded health data and the challenge of interpreting it adequately. Further, due to remaining challenges, such as sensor stability, sample standardization, and detecting low biomarker concentrations, an easy, lab-independent method of detecting complex molecules (eg, proteins) remains an unmet target and requires substantial development and validation efforts. Smart biosensors represent a great potential to solve remaining health care challenges in high-, middle-, and low-income countries. For successful clinical implementation, it will be essential to bring together all stakeholders such as engineers, health care workers, and business partners early, to enable a cost-effective and valuable solution.

Novel Biomarkers in Sweat, Saliva, and Exhaled Breath to Detect Active Tuberculosis

aTbc is caused by infection with a member of the *Mycobacterium tuberculosis* complex. aTbc is endemic in India, Indonesia, China, Nigeria, Pakistan, and South Africa. From 2000 to 2016, an estimated 53 million people died from tuberculosis infection, and about 10.4 million people were infected per year [20]. To fulfil the WHO End TB Strategy, novel diagnostic markers and devices are urgently needed, as a reliable, straightforward, fast diagnostic test is still lacking [37,38]. Present diagnostic tests are mostly sputum-dependent, which is a significant limitation given that many patients do not produce sputum [39-41]. If sputum cannot be collected spontaneously, an invasive bronchoalveolar lavage is indicated to confirm the diagnosis. Due to limitations of current tests, the patient may remain in contact isolation for weeks until diagnosis. A straightforward test to detect aTbc would enable earlier diagnosis and treatment initiation and therefore potentially lead to a lower transmission rate and decreased morbidity, mortality, and health care costs [42-44].

Biomarkers in Sweat

By using mass spectrometry analysis, Adewole et al [45] detected 26 specific proteins in the sweat of patients diagnosed with aTbc (Table 1). The researchers were able to differentiate between patients with aTbc, patients with a differential diagnosis such as pneumonia or lung cancer, and healthy controls by taking into account the protein expressions in sweat. Differentiation between the groups was even possible for patients with a history of adequately treated aTbc. The proteomic sweat analysis revealed specific molecules for aTbc such as C1q subcomponent, which has been associated with aTbc [46].

Table 1. Overview of next-generation digital biomarkers in sweat, saliva, and exhaled breath for infectious diseases.

Disease or biofluid	In sweat	In saliva	In exhaled breath
Active tuberculosis	Complement C1q subcomponent subunit C, C-reactive protein [45]	Interleukin-1β, interleukin-23, ECM-1, HCC-1, and fibrinogen [47]	Oxetane, 3-(1-methyethyl)-, dodecane, 4-methyl-, cyclohexane, hexyl- [48]
Therapeutic antibiotic drug monitoring	Fluorquinolone (ciprofloxacin), cephalosporines (eg, ceftriaxone) [49]	Cephalosporines (eg, cefuroxime), macrolides (eg, clarithromycin), tetracycline (eg, doxicycline), fluorquinolone (eg, ciprofloxacine) [50]	Carbapeneme (meropenem), acylaminopenicillin (piperacillin) [51]

Biomarkers in Saliva

Jacobs et al [47] investigated saliva samples from patients presenting with suspected aTbc prior to diagnosis. Patients were subsequently classified using clinical, radiological, and laboratory findings as having aTBc or other respiratory diseases. A 5-biomarker signature detected by Luminex multiplex immunoassay including markers such as interleukin-1β (Table 1) enabled detection of aTbc with a sensitivity of 88.90% and a specificity of 89.75%. Extending the biomarker range led to a sensitivity of 100% and a specificity of 95% (in the absence of HIV) and detected a treatment response from the host after adequate treatment initiation [47].

Biomarkers in Exhaled Breath

In 1998, Wang et al [52] investigated patients with aTbc and described different levels of nitric oxide in exhaled breath as a potential diagnostic marker using chemiluminescence analysis. By gas chromatographic and mass spectrometric analysis, Phillips et al [48] detected further promising volatile components to differentiate between aTbc and controls (Table 1). Further Nakhleh et al [53] investigated 198 samples from aTbc patients and healthy controls. Of them, 138 samples with known diagnosis were used for unsupervised sensor training, and 60 undefined samples were used for blind validation. In addition, combining exhaled analysis by nanomaterial-based sensors with unsupervised machine learning led to a promising, noninvasive detection tool [54]. Further investigation supported this approach, by using “breathprints” for the detection of *M. tuberculosis* [55].

Perspective on Digital Biomarkers to Detect aTbc

Detecting and diagnosing aTbc remain laborious and time consuming. Increasing resistance against antimicrobial therapies is further aggravating the burden caused by the disease. As aTbc is mainly endemic in low- and middle-income countries with rather restricted health care access, novel approaches to detect, diagnose, and treat aTbc are needed. An increasing amount of basic research has been conducted recently to define novel aTbc biomarkers. Next to well-known proteins such as C1q and CRP, novel and promising proteins have been explored in sweat, saliva, and exhaled breath of aTbc patients that need further confirmation of value [45]. Smart biosensors for detecting specific aTbc molecules in the aforementioned biofluids have not been applied yet, while e-NOSE has to be considered as an exception [56,57]. Therefore, exhaled breath is the most advanced biofluid when it comes to clinical implementation, while sweat and saliva remain very promising concepts. Nevertheless, exhaled breath may be restricted to pulmonary tuberculosis only. Therefore, sweat and saliva analysis may provide deeper systematic insights into not only pulmonary but also extrapulmonary aTbc.

Detecting Antibiotics in Sweat, Saliva, and Exhaled Breath for Therapeutic Drug Monitoring

Increasing antimicrobial resistance (AMR) is one of the biggest threats to global health, food security, and development. AMR has a high impact not only on prevention but also on treatment of infectious diseases [58]. Antibiotic misuse and overuse are the main accelerators of the development of AMR [59]. Approximately >30,000 deaths, >870,000 disability-adjusted life years, and >670,000 infections were caused by multidrug resistant organisms in the European Union in 2015 [60]. Antibiotic-resistant bacterial infections lead to a higher rate of complications and a significantly higher need for resources [59]. Improved sanitation, hospital hygiene, infection precautions, and antimicrobial stewardship are proposed solutions to reduce AMR. An important aspect of antimicrobial stewardship is therapeutic drug monitoring (TDM), which is already the standard of care for the treatment with glycopeptides because of their toxicity and TDM’s positive effects on patients’ clinical outcomes [61]. Adult dose recommendations for antibiotics account only barely for renal function, liver function, and weight. Appropriate dosing is crucial, as previous evidence indicates significant interpersonal variability in antibiotic serum concentration and a low chance of achieving even the most conservative pharmacological target [62]. Despite the advances in TDM, a reliable, real-time, noninvasive, and lab-independent monitoring approach is still not available [63,64].

Antibiotics in Sweat

Høiby et al [49,65] detected ciprofloxacin in sweat in 1997 and certain β-lactam antibiotics in a follow-up study (Table 1) using a microbiological agar diffusion method. Increasing amounts of ciprofloxacin in axillary sweat were associated with the length of antibiotic application time and remained detectable up to 28 hours after the last intake. Further, antibiotic concentrations in sweat may even qualify as a surrogate marker for tissue penetration in soft tissue infections as demonstrated by Brasier et al [66] using ultra high-performance liquid chromatography coupled with quadrupole mass spectrometry. Larger studies that correlate antibiotic concentrations in sweat with clinical outcome are needed to prove this concept.

Antibiotics in Saliva

Different classes of antibiotics have been detected in saliva so far [50]. Intravenously administered cephalosporins as well as orally administered cephalosporins, tetracycline, and fluoroquinolones are excreted in concentrations above the minimal inhibitory concentration (MIC) in sialadenitis (Table 1) [50,67,68]. However, detected by microbiological agar diffusion, among other methods, phenoxymethylpenicillin was shown to be secreted only in very low concentrations and was therefore deemed not a reasonable treatment approach for sialadenitis [50,67]. Macrolides are an additional class of antibiotics that have been detected in saliva, but in the aforementioned case of sialadenitis, their use has not been indicated due to a high probability of developing antibiotic resistances [69].

Antibiotics in Exhaled Breath

Herregodts et al [51] recruited 11 critically ill patients from the intensive care unit to investigate the detectability of piperacillin and meropenem in exhaled breath using ultra high-pressure liquid chromatography high-resolution mass spectrometry (Table 1). After collecting breath using the ExaBreath, breath samples were analyzed by mass spectrometry. Piperacillin was detected at a median of 3083 pg/filter (988-203,895 pg/filter) and meropenem at 21,168 pg/filter. However, antibiotic concentrations of 2 of 11 patients were below the lower limit of quantification. Further, antibiotic concentrations in exhaled breath and blood did not correlate. The authors finally concluded that antibiotics in exhaled breath potentially represent the epithelial lining fluid concentrations. More investigations are needed to better understand lung tissue concentrations rather than comparing concentrations in exhaled breath to plasma [51].

Perspective on Digital Biomarkers for TDM

For sweat analysis, it is unclear if antibiotic concentrations in sweat correlate with the concentrations in blood. However, sweat concentrations may serve as a surrogate marker for antibiotic tissue penetration. This may allow for targeting the detected bacterial MIC obtained by microbial tissue samples in the tissue [65,70]. TDM in saliva potentially shows a comparable weakness, as correlating antibiotic concentrations in saliva and blood remains too ambitious of a target. Still, monitoring antibiotic concentrations in saliva may allow for ensuring antibiotic concentrations above the MIC in bacterial sialadenitis and provide an achievable target in the short term [50]. For exhaled breath, lung diseases such as bacterial pneumonia or pulmonary tuberculosis with a defined MIC seem to be a further achievable monitoring approach.

Overall, it remains to be determined if antibiotic concentrations in sweat, saliva, and exhaled breath correlate with antibiotic concentrations in blood and may serve as systemic, noninvasive TDM in the future. It is highly intuitive that local antibiotic concentrations can potentially be used as a surrogate marker and support the adequate administration of antibiotic drugs in a foreseeable timeframe.

Discussion

Principal Findings

From this perspective, we discussed the potential application of novel smart biosensors in aTbc and antimicrobial multidrug resistance, 2 urgent global issues in infectious diseases with unmet needs. Emerging smart biosensors analyzing sweat, saliva, and exhaled breath are increasingly available. These biosensors enable noninvasive sample collection and provide a remote, lab-independent patient monitoring system if combined with smartphones. Moreover, integrating artificial intelligence into the analysis process is a promising addition as it has the potential to support significantly better disease detection and discrimination [71]. Different sensor approaches for on-skin sweat analysis, tooth-mounted sensors for saliva analysis, and e-Noses for exhaled breath analysis are under development at present [25,26,29]. Several specific molecular markers such as the C1q subunit in sweat and interleukin-1β in saliva were detected in patients with aTbc [45,47], thus opening up a new pool of noninvasive biomarkers. Further, different antimicrobials such as β-lactam antibiotics or tetracycline have been detected in sweat, saliva, or exhaled breath [51,65,72]. AMR is a major global threat, and antibiotic stewardship is a proven approach to reduce the development of bacterial resistance [58,73]. Despite the promising approaches and concepts that are considered for leveraging disease burden, only the e-NOSE has been clinically implemented and has to be seen as an exception in this explorative field [56].

A noninvasive, decentralized, lab-independent tool to monitor health and disease is a promising approach to optimize patient care in the future. However, major challenges, such as the lack of technology readiness levels of biosensors, unknown correlations, and high variability between concentrations in sweat, saliva, exhaled breath, and blood, need to be addressed. Before clinical investigation and implementation, a strict and structured validation in laboratory settings is needed to provide internal biosensor validity and clinically interpretable data. Biosensors have to work in different environmental settings, especially when it comes to endemic areas for aTbc, but for

TDM as well. Heretofore, high temperatures and high humidity are main challenges to be addressed concerning sensor stability. While biosensor development is expensive and raising funding remains challenging, intensifying basic clinical research and demonstrating clinical value are key to accelerating targeted biosensor development and validation. Due to the decentralized, lab-independent approach and ubiquitous application, smart biosensors provide a scalable approach to impact the health care of the future.

Further Limitations

Despite the high potential, a few issues remain and have been keeping biosensors from broad clinical implementation, such as low biomarker concentrations and unstable samples, as discussed earlier. In addition, a variety of methods ranging from microbiological agar diffusion methods to mass spectrometry have been used for biofluid analysis. It remains to be determined if the heterogeneously analyzed biomarkers will be detectable by different smart biosensors, and challenges in biofluid sampling need to be addressed during device development. There are further challenging factors for each biofluid.

Sweat

Electrolyte concentrations in sweat are known to be dynamic over time when exercising, and sweat rate is known to change with alterations in body temperature [74]. Further, differences in age, gender, and ethnicity have been described and need to be further investigated [75].

Saliva

Several physiological factors influence saliva composition; eating habits as well as oral health have also demonstrated a significant impact [76,77]. Moreover, saliva can be contaminated with blood, leading to concentration changes in oxidative stress markers, for example, such as after brushing teeth [78].

Exhaled Breath

Detection of biomarkers in exhaled breath is influenced by different habits. Diet and oral health have proven to challenge breath analytics by varying the components [77,79]. Smoking leads to further changes in the breath pattern [80].

Conclusion

We see great potential in applying smart biosensors to the detection of next-generation digital biomarkers in infectious diseases and provide the first noninvasive, lab-independent, digital point-of-care diagnostics in the future. Modularly combining sweat, saliva, and breath sensors with additional biosensors to collect accelerometer data, heart rate, and body temperature will enable more context-driven monitoring and will support sample standardization. Moreover, if combined with a GPS tracker, those devices allow instant interpretation and monitoring of the spread of infections independently of health care infrastructure on a global scale. To fully investigate the potential of smartphone-based biosensors, more effort is needed to develop, validate, and make those biosensors globally available.

Authors' Contributions

NB contributed substantially to the conception of the study and analysis of the state-of-the-art literature. NB, MO, FDI, and JE drafted the work and revised it critically for important intellectual content. NB, MO, FDI, and JE approved the version to be published and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Conflicts of Interest

JE holds 0.5% virtual shares of Preventicus GmbH.

References

1. Biomarkers Definitions Working Group.. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther 2001 Mar;69(3):89-95. [doi: 10.1067/mcp.2001.113989] [Medline: 11240971]
2. Yoon C, Semitala FC, Atuhumuza E, Katende J, Mwebe S, Asege L, et al. Point-of-care C-reactive protein-based tuberculosis screening for people living with HIV: a diagnostic accuracy study. The Lancet Infectious Diseases 2017 Dec;17(12):1285-1292. [doi: 10.1016/s1473-3099(17)30488-7]
3. Pujades-Rodriguez M, West RM, Wilcox MH, Sandoe J. Lower Urinary Tract Infections: Management, Outcomes and Risk Factors for Antibiotic Re-prescription in Primary Care. EClinicalMedicine 2019 Sep;14:23-31 [FREE Full text] [doi: 10.1016/j.eclinm.2019.07.012] [Medline: 31709399]
4. Firouzi F, Rahmani AM, Mankodiya K, Badaroglu M, Merrett G, Wong P, et al. Internet-of-Things and big data for smarter healthcare: From device to architecture, applications and analytics. Future Generation Computer Systems 2018 Jan;78:583-586. [doi: 10.1016/j.future.2017.09.016]
5. Babrak L, Menetski J, Rebhan M, Nisato G, Zinggeler M, Brasier N, et al. Traditional and Digital Biomarkers: Two Worlds Apart? Digit Biomark 2019;3(2):92-102 [FREE Full text] [doi: 10.1159/000502000] [Medline: 32095769]
6. Malasinghe LP, Ramzan N, Dahal K. Remote patient monitoring: a comprehensive study. J Ambient Intell Human Comput 2017 Oct 26;10(1):57-76. [doi: 10.1007/s12652-017-0598-x]
7. Tom-Aba D, Nguku PM, Arinze CC, Krause G. Assessing the Concepts and Designs of 58 Mobile Apps for the Management of the 2014-2015 West Africa Ebola Outbreak: Systematic Review. JMIR Public Health Surveill 2018 Oct 29;4(4):e68 [FREE Full text] [doi: 10.2196/publichealth.9015] [Medline: 30373727]

8. Wyss T, Roos L, Beeler N, Veenstra B, Delves S, Buller M, et al. The comfort, acceptability and accuracy of energy expenditure estimation from wearable ambulatory physical activity monitoring systems in soldiers. Journal of Science and Medicine in Sport 2017 Nov;20:S133-S134. [doi: 10.1016/j.jsams.2017.09.498]
9. Weikert M, Motl RW, Suh Y, McAuley E, Wynn D. Accelerometry in persons with multiple sclerosis: measurement of physical activity or walking mobility? J Neurol Sci 2010 Mar 15;290(1-2):6-11. [doi: 10.1016/j.jns.2009.12.021] [Medline: 20060544]
10. Rovini E, Maremmani C, Cavallo F. How Wearable Sensors Can Support Parkinson's Disease Diagnosis and Treatment: A Systematic Review. Front Neurosci 2017;11:555 [FREE Full text] [doi: 10.3389/fnins.2017.00555] [Medline: 29056899]
11. Brasier N, Raichle CJ, Dörr M, Becke A, Nohturfft V, Weber S, et al. Detection of atrial fibrillation with a smartphone camera: first prospective, international, two-centre, clinical validation study (DETECT AF PRO). Europace 2019 Jan 01;21(1):41-47 [FREE Full text] [doi: 10.1093/europace/euy176] [Medline: 30085018]
12. Chen H, Liu K, Li Z, Wang P. Point of care testing for infectious diseases. Clin Chim Acta 2019 Jun;493:138-147 [FREE Full text] [doi: 10.1016/j.cca.2019.03.008] [Medline: 30853460]
13. Kosack CS, Page A, Klatser PR. A guide to aid the selection of diagnostic tests. Bull. World Health Organ 2017 Jun 26;95(9):639-645. [doi: 10.2471/blt.16.187468]
14. Yager P, Domingo GJ, Gerdes J. Point-of-care diagnostics for global health. Annu Rev Biomed Eng 2008;10:107-144. [doi: 10.1146/annurev.bioeng.10.061807.160524] [Medline: 18358075]
15. Fogel R, Limson J. Developing Biosensors in Developing Countries: South Africa as a Case Study. Biosensors (Basel) 2016 Feb 02;6(1):5 [FREE Full text] [doi: 10.3390/bios6010005] [Medline: 26848700]
16. Parkes-Ratanshi R, Kikonyogo R, Hsieh Y, Nakku-Joloba E, Manabe YC, Gaydos CA, et al. Point-of-care diagnostics: needs of African health care workers and their role combating global antimicrobial resistance. Int J STD AIDS 2019 Mar;30(4):404-410 [FREE Full text] [doi: 10.1177/0956462418807112] [Medline: 30626282]
17. Kuupiel D, Bawontuo V, Mashamba-Thompson T. Improving the Accessibility and Efficiency of Point-of-Care Diagnostics Services in Low- and Middle-Income Countries: Lean and Agile Supply Chain Management. Diagnostics (Basel) 2017 Nov 29;7(4):58 [FREE Full text] [doi: 10.3390/diagnostics7040058] [Medline: 29186013]
18. Ates HC, Roberts JA, Lipman J, Cass AE, Urban GA, Dincer C. On-Site Therapeutic Drug Monitoring. Trends Biotechnol 2020 Nov;38(11):1262-1277 [FREE Full text] [doi: 10.1016/j.tibtech.2020.03.001] [Medline: 33058758]
19. Brasier N, Eckstein J. Sweat as a Source of Next-Generation Digital Biomarkers. Digit Biomark 2019;3(3):155-165 [FREE Full text] [doi: 10.1159/000504387] [Medline: 32095774]
20. Global Tuberculosis Report. World Health Organization. 2018. URL: https://apps.who.int/iris/bitstream/handle/10665/274453/9789241565646-eng.pdf [accessed 2021-07-04]
21. Ferri M, Ranucci E, Romagnoli P, Giaccone V. Antimicrobial resistance: A global emerging threat to public health systems. Crit Rev Food Sci Nutr 2017 Sep 02;57(13):2857-2876. [doi: 10.1080/10408398.2015.1077192] [Medline: 26464037]
22. Yu Y, Prassas I, Muytjens CM, Diamandis EP. Proteomic and peptidomic analysis of human sweat with emphasis on proteolysis. J Proteomics 2017 Feb 23;155:40-48. [doi: 10.1016/j.jprot.2017.01.005] [Medline: 28095327]
23. van der Schee MP, Paff T, Brinkman P, van Aalderen WMC, Haarman E, Sterk P. Breathomics in lung disease. Chest 2015 Jan;147(1):224-231 [FREE Full text] [doi: 10.1378/chest.14-0781] [Medline: 25560860]
24. Hu S, Xie Y, Ramachandran P, Ogorzalek Loo RR, Li Y, Loo JA, et al. Large-scale identification of proteins in human salivary proteome by liquid chromatography/mass spectrometry and two-dimensional gel electrophoresis-mass spectrometry. Proteomics 2005 Apr;5(6):1714-1728. [doi: 10.1002/pmic.200401037] [Medline: 15800970]
25. Tseng P, Napier B, Garbarini L, Kaplan DL, Omenetto FG. Functional, RF-Trilayer Sensors for Tooth-Mounted, Wireless Monitoring of the Oral Cavity and Food Consumption. Adv Mater 2018 May;30(18):e1703257. [doi: 10.1002/adma.201703257] [Medline: 29572979]
26. Karakaya D, Ulucan O, Turkan M. Electronic Nose and Its Applications: A Survey. Int. J. Autom. Comput 2019 Dec 28;17(2):179-209. [doi: 10.1007/s11633-019-1212-9]
27. Reeder JT, Choi J, Xue Y, Gutruf P, Hanson J, Liu M, et al. Waterproof, electronics-enabled, epidermal microfluidic devices for sweat collection, biomarker analysis, and thermography in aquatic settings. Sci Adv 2019 Jan;5(1):eaau6356 [FREE Full text] [doi: 10.1126/sciadv.aau6356] [Medline: 30746456]
28. Gao W, Emaminejad S, Nyein HYY, Challa S, Chen K, Peck A, et al. Fully integrated wearable sensor arrays for multiplexed in situ perspiration analysis. Nature 2016 Jan 28;529(7587):509-514 [FREE Full text] [doi: 10.1038/nature16521] [Medline: 26819044]
29. Orro K, Smirnova O, Arshavskaja J, Salk K, Meikas A, Pihelgas S, et al. Development of TAP, a non-invasive test for qualitative and quantitative measurements of biomarkers from the skin surface. Biomark Res 2014;2(1):20 [FREE Full text] [doi: 10.1186/2050-7771-2-20] [Medline: 25785188]
30. Dalirirad S, Steckl A. Aptamer-based lateral flow assay for point of care cortisol detection in sweat. Sensors and Actuators B: Chemical 2019 Mar;283:79-86 [FREE Full text] [doi: 10.1016/j.snb.2018.11.161]
31. Yang Y, Song Y, Bo X, Min J, Pak OS, Zhu L, et al. A laser-engraved wearable sensor for sensitive detection of uric acid and tyrosine in sweat. Nat Biotechnol 2020 Feb;38(2):217-224. [doi: 10.1038/s41587-019-0321-x] [Medline: 31768044]
32. The eNose Company. URL: https://www.enose.nl/products/aeonose/ [accessed 2021-07-04]

33. Hashoul D, Haick H. Sensors for detecting pulmonary diseases from exhaled breath. Eur Respir Rev 2019 Jun 30;28(152):190011 [FREE Full text] [doi: 10.1183/16000617.0011-2019] [Medline: 31243097]
34. Dhar P, Kashyap P, Jindal N, Rani R. Role of electronic nose technology in food industry. 2018 Presented at: Emerging sustainable technologies in food processing (ESTFP-2018); March 15-16, 2018; Sangrur, India.
35. Sabatini A, Zompanti A, Grasso S, Gianfelici A, Di Castro A, Donatucci B, et al. Design And Development Of An Innovative Sensor System For Non-Invasive Monitoring Of Athletic Performances. 2019 Presented at: II Workshop on Metrology for Industry 4.0 and IoT (MetroInd4.0&IoT); June 4-6, 2019; Naples, Italy. [doi: 10.1109/metroi4.2019.8792863]
36. Mohamed EI, Mohamed MA, Moustafa MH, Abdel-Mageed SM, Moro AM, Baess AI, et al. Qualitative analysis of biological tuberculosis samples by an electronic nose-based artificial neural network. Int J Tuberc Lung Dis 2017 Jul 01;21(7):810-817. [doi: 10.5588/ijtld.16.0677] [Medline: 28633707]
37. Global Tuberculosis Report. World Health Organization. 2017. URL: https://apps.who.int/iris/bitstream/handle/10665/259366/9789241565516-eng.pdf?sequence=1 [accessed 2021-07-04]
38. Boehme CC, Nicol MP, Nabeta P, Michael JS, Gotuzzo E, Tahirli R, et al. Feasibility, diagnostic accuracy, and effectiveness of decentralised use of the Xpert MTB/RIF test for diagnosis of tuberculosis and multidrug resistance: a multicentre implementation study. The Lancet 2011 Apr;377(9776):1495-1505. [doi: 10.1016/S0140-6736(11)60438-8]
39. Sharma S, Mohan A. Extrapulmonary tuberculosis. Indian J Med Res 2004 Oct;120(4):316-353. [Medline: 15520485]
40. Bacakoğlu F, Başoğlu, Cok G, Sayiner A, Ateş M. Pulmonary tuberculosis in patients with diabetes mellitus. Respiration 2001;68(6):595-600. [doi: 10.1159/000050578] [Medline: 11786714]
41. Carrol E, Clark J, Cant A. Non-pulmonary tuberculosis. Paediatr Respir Rev 2001 Jun;2(2):113-119. [doi: 10.1053/prrv.2000.0118] [Medline: 12531057]
42. Millen SJ, Uys PW, Hargrove J, van Helden PD, Williams BG. The effect of diagnostic delays on the drop-out rate and the total delay to diagnosis of tuberculosis. PLoS One 2008 Apr 09;3(4):e1933 [FREE Full text] [doi: 10.1371/journal.pone.0001933] [Medline: 18398459]
43. Havlir DV, Barnes PF. Tuberculosis in Patients with Human Immunodeficiency Virus Infection. N Engl J Med 1999 Feb 04;340(5):367-373. [doi: 10.1056/nejm199902043400507]
44. Roth JA, Hornung-Winter C, Radicke I, Hug BL, Biedert M, Abshagen C, et al. Direct Costs of a Contact Isolation Day: A Prospective Cost Analysis at a Swiss University Hospital. Infect Control Hosp Epidemiol 2018 Jan;39(1):101-103. [doi: 10.1017/ice.2017.258] [Medline: 29249218]
45. Adewole OO, Erhabor GE, Adewole TO, Ojo AO, Oshokoya H, Wolfe LM, et al. Proteomic profiling of eccrine sweat reveals its potential as a diagnostic biofluid for active tuberculosis. Proteomics Clin Appl 2016 May;10(5):547-553. [doi: 10.1002/prca.201500071] [Medline: 26948146]
46. Cai Y, Yang Q, Tang Y, Zhang M, Liu H, Zhang G, et al. Increased complement C1q level marks active disease in human tuberculosis. PLoS One 2014;9(3):e92340 [FREE Full text] [doi: 10.1371/journal.pone.0092340] [Medline: 24647646]
47. Jacobs R, Maasdorp E, Malherbe S, Loxton AG, Stanley K, van der Spuy G, et al. Diagnostic Potential of Novel Salivary Host Biomarkers as Candidates for the Immunological Diagnosis of Tuberculosis Disease and Monitoring of Tuberculosis Treatment Response. PLoS One 2016 Aug 3;11(8):e0160546 [FREE Full text] [doi: 10.1371/journal.pone.0160546] [Medline: 27487181]
48. Phillips M, Basa-Dalay V, Bothamley G, Cataneo RN, Lam PK, Natividad MPR, et al. Breath biomarkers of active pulmonary tuberculosis. Tuberculosis (Edinb) 2010 Mar;90(2):145-151. [doi: 10.1016/j.tube.2010.01.003] [Medline: 20189456]
49. Høiby N, Jarløv JO, Kemp M, Tvede M, Bangsborg JM, Kjerulf A, et al. Excretion of ciprofloxacin in sweat and multiresistant Staphylococcus epidermidis. The Lancet 1997 Jan;349(9046):167-169. [doi: 10.1016/s0140-6736(96)09229-x]
50. Troeltzsch M, Pache C, Probst F, Troeltzsch M, Ehrenfeld M, Otto S. Antibiotic concentrations in saliva: a systematic review of the literature, with clinical implications for the treatment of sialadenitis. J Oral Maxillofac Surg 2014 Jan;72(1):67-75. [doi: 10.1016/j.joms.2013.06.214] [Medline: 23992785]
51. Herregodts J, Van Vooren S, Deschuyteneer E, Dhaese SAM, Stove V, Verstraete AG, et al. Measuring antibiotics in exhaled air in critically ill, non-ventilated patients: A feasibility and proof of concept study. J Crit Care 2019 Jun;51:46-50. [doi: 10.1016/j.jcrc.2019.01.025] [Medline: 30745285]
52. Wang C, Liu C, Lin H, Yu C, Chung K, Kuo H. Increased exhaled nitric oxide in active pulmonary tuberculosis due to inducible NO synthase upregulation in alveolar macrophages. Eur Respir J 1998 Apr;11(4):809-815 [FREE Full text] [doi: 10.1183/09031936.98.11040809] [Medline: 9623681]
53. Nakhleh MK, Jeries R, Gharra A, Binder A, Broza YY, Pascoe M, et al. Detecting active pulmonary tuberculosis with a breath test using nanomaterial-based sensors. Eur Respir J 2014 May;43(5):1522-1525 [FREE Full text] [doi: 10.1183/09031936.00019114] [Medline: 24789953]
54. Beccaria M, Bobak C, Maitshotlo B, Mellors TR, Purcaro G, Franchina FA, et al. Exhaled human breath analysis in active pulmonary tuberculosis diagnostics by comprehensive gas chromatography-mass spectrometry and chemometric techniques. J Breath Res 2018 Nov 05;13(1):016005 [FREE Full text] [doi: 10.1088/1752-7163/aae80e] [Medline: 30394364]
55. Zetola N, Modongo C, Matsiri O, Tamuhla T, Mbongwe B, Matlhagela K, et al. Diagnosis of pulmonary tuberculosis and assessment of treatment response through analyses of volatile compound patterns in exhaled breath samples. J Infect 2017 Apr;74(4):367-376 [FREE Full text] [doi: 10.1016/j.jinf.2016.12.006] [Medline: 28017825]

56. Bruins M, Rahim Z, Bos A, van de Sande WW, Endtz HP, van Belkum A. Diagnosis of active tuberculosis by e-nose analysis of exhaled air. Tuberculosis (Edinb) 2013 Mar;93(2):232-238. [doi: 10.1016/j.tube.2012.10.002] [Medline: 23127779]
57. Saktiawati AMI, Stienstra Y, Subronto YW, Rintiswati N, Sumardi, Gerritsen J, et al. Sensitivity and specificity of an electronic nose in diagnosing pulmonary tuberculosis among patients with suspected tuberculosis. PLoS One 2019;14(6):e0217963 [FREE Full text] [doi: 10.1371/journal.pone.0217963] [Medline: 31194793]
58. Global action plan on antimicrobial resistance. World Health Organization. 2016 Jan 01. URL: https://www.who.int/publications/i/item/9789241509763 [accessed 2021-07-04]
59. Antibiotic resistance: Key facts. World Health Organization. 2020 Jul 31. URL: https://www.who.int/news-room/fact-sheets/detail/antibiotic-resistance [accessed 2021-07-04]
60. Cassini A, Högberg LD, Plachouras D, Quattrocchi A, Hoxha A, Simonsen GS, et al. Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis. The Lancet Infectious Diseases 2019 Jan;19(1):56-66. [doi: 10.1016/S1473-3099(18)30605-4]
61. Ye Z, Tang H, Zhai S. Benefits of therapeutic drug monitoring of vancomycin: a systematic review and meta-analysis. PLoS One 2013;8(10):e77169 [FREE Full text] [doi: 10.1371/journal.pone.0077169] [Medline: 24204764]
62. Roberts JA, Paul SK, Akova M, Bassetti M, De Waele JJ, Dimopoulos G, DALI Study. DALI: defining antibiotic levels in intensive care unit patients: are current β-lactam antibiotic doses sufficient for critically ill patients? Clin Infect Dis 2014 Apr;58(8):1072-1083. [doi: 10.1093/cid/ciu027] [Medline: 24429437]
63. Carlier M, Stove V, Wallis S, De Waele JJ, Verstraete AG, Lipman J, et al. Assays for therapeutic drug monitoring of β-lactam antibiotics: A structured review. Int J Antimicrob Agents 2015 Oct;46(4):367-375. [doi: 10.1016/j.ijantimicag.2015.06.016] [Medline: 26271599]
64. Sime FB, Roberts MS, Tiong IS, Gardner JH, Lehman S, Peake SL, et al. Can therapeutic drug monitoring optimize exposure to piperacillin in febrile neutropenic patients with haematological malignancies? A randomized controlled trial. J Antimicrob Chemother 2015 Aug;70(8):2369-2375. [doi: 10.1093/jac/dkv123] [Medline: 25953805]
65. Høiby N, Pers C, Johansen HK, Hansen H. Excretion of β-Lactam Antibiotics in Sweat—a Neglected Mechanism for Development of Antibiotic Resistance? Antimicrob Agents Chemother 2000 Oct 01;44(10):2855-2857. [doi: 10.1128/aac.44.10.2855-2857.2000]
66. Brasier N, Widmer A, Osthoff M, Mutke M, De Ieso F, Brasier-Lutz P, et al. Non-invasive Drug Monitoring of β-Lactam Antibiotics Using Sweat Analysis-A Pilot Study. Front Med (Lausanne) 2020;7:476 [FREE Full text] [doi: 10.3389/fmed.2020.00476] [Medline: 32984371]
67. Strömberg A, Friberg U, Cars O. Concentrations of phenoxymethylpenicillin and cefadroxil in tonsillar tissue and tonsillar surface fluid. Eur. J, Clin. Microbiol 1987 Oct;6(5):525-529. [doi: 10.1007/bf02014240]
68. Najjar T, Alkharfy K, Saad S. Mechanism and implication of cephalosporin penetration into oropharyngeal mucosa. J Infect Chemother 2009 Apr;15(2):70-74. [doi: 10.1007/s10156-008-0666-4] [Medline: 19396515]
69. Rams T, Dujardin S, Sautter J, Degener J, van Winkelhoff AJ. Spiramycin resistance in human periodontitis microbiota. Anaerobe 2011 Aug;17(4):201-205. [doi: 10.1016/j.anaerobe.2011.03.017] [Medline: 21524712]
70. Kazimoto T, Abdulla S, Bategereza L, Juma O, Mhimbira F, Weisser M, et al. Causative agents and antimicrobial resistance patterns of human skin and soft tissue infections in Bagamoyo, Tanzania. Acta Trop 2018 Oct;186:102-106. [doi: 10.1016/j.actatropica.2018.07.007] [Medline: 30006029]
71. Nakhleh M, Amal H, Jeries R, Broza YY, Aboud M, Gharra A, et al. Diagnosis and Classification of 17 Diseases from 1404 Subjects via Pattern Analysis of Exhaled Molecules. ACS Nano 2017 Jan 24;11(1):112-125 [FREE Full text] [doi: 10.1021/acsnano.6b04930] [Medline: 28000444]
72. Quayle A, Hailey D. Antimicrobial substances in saliva: sulphamethoxazole and trimethoprim (co-trimoxazole). British Journal of Oral Surgery 1973 Jul;11(1):60-65 [FREE Full text] [doi: 10.1016/0007-117x(73)90077-2]
73. Roberts J, Norris R, Paterson D, Martin J. Therapeutic drug monitoring of antimicrobials. Br J Clin Pharmacol 2012 Jan;73(1):27-36 [FREE Full text] [doi: 10.1111/j.1365-2125.2011.04080.x] [Medline: 21831196]
74. Baker LB. Sweating Rate and Sweat Sodium Concentration in Athletes: A Review of Methodology and Intra/Interindividual Variability. Sports Med 2017 Mar;47(Suppl 1):111-128 [FREE Full text] [doi: 10.1007/s40279-017-0691-5] [Medline: 28332116]
75. Lee J, Kim J, Murota H. Perspiration Functions in Different Ethnic, Age, and Sex Populations: Modification of Sudomotor Function. Curr Probl Dermatol 2016;51:109-119. [doi: 10.1159/000447370] [Medline: 27584970]
76. Johansson A, Norring C, Unell L, Johansson A. Eating disorders and biochemical composition of saliva: a retrospective matched case-control study. Eur J Oral Sci 2015 Jun;123(3):158-164 [FREE Full text] [doi: 10.1111/eos.12179] [Medline: 25780814]
77. van den Broek AM, Feenstra L, de Baat C. A review of the current literature on aetiology and measurement methods of halitosis. J Dent 2007 Aug;35(8):627-635. [doi: 10.1016/j.jdent.2007.04.009] [Medline: 17555859]

78. Kamodyová N, Baňasová L, Janšáková K, Koborová I, Tóthová Ľ, Stanko P, et al. Blood Contamination in Saliva: Impact on the Measurement of Salivary Oxidative Stress Markers. Dis Markers 2015;2015:479251 [FREE Full text] [doi: 10.1155/2015/479251] [Medline: 26347580]
79. Ajibola OA, Smith D, Spaněl P, Ferns GAA. Effects of dietary nutrients on volatile breath metabolites. J Nutr Sci 2013;2:e34 [FREE Full text] [doi: 10.1017/jns.2013.26] [Medline: 25191584]
80. Capone S, Tufariello M, Forleo A, Longo V, Giampetruzzi L, Radogna AV, et al. Chromatographic analysis of VOC patterns in exhaled breath from smokers and nonsmokers. Biomed Chromatogr 2018 Apr;32(4):e1001940 [FREE Full text] [doi: 10.1002/bmc.4132] [Medline: 29131420]

Abbreviations

AMR: antimicrobial resistance
aTbc: active tuberculosis
CRP: C-reactive protein
ELISA: enzyme-linked immunoassay
EN: electric nose
MIC: minimal inhibitory concentration
POCT: point-of-care testing
TDM: therapeutic drug monitoring
WHO: World Health Organization

Edited by R Kukafka; submitted 20.11.20; peer-reviewed by CY Yeh, V Shah, I Tarnanas; comments to author 15.01.21; revised version received 25.01.21; accepted 24.05.21; published 19.08.21.

Please cite as:
Brasier N, Osthoff M, De Ieso F, Eckstein J
Next-Generation Digital Biomarkers for Tuberculosis and Antibiotic Stewardship: Perspective on Novel Molecular Digital Biomarkers in Sweat, Saliva, and Exhaled Breath
J Med Internet Res 2021;23(8):e25907
URL: https://www.jmir.org/2021/8/e25907
doi:10.2196/25907
PMID:34420925

©Noe Brasier, Michael Osthoff, Fiorangelo De Ieso, Jens Eckstein. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Classification of Children With Autism and Typical Development Using Eye-Tracking Data From Face-to-Face Conversations: Machine Learning Model Development and Performance Evaluation

Zhong Zhao[1], PhD; Haiming Tang[1]; Xiaobin Zhang[2], PhD; Xingda Qu[1], PhD; Xinyao Hu[1], PhD; Jianping Lu[3]

[1]Institute of Human Factors and Ergonomics, College of Mechatronics and Control Engineering, Shenzhen University, Shenzhen, China
[2]Shenzhen Guangming District Center for Disease Control and Prevention, Shenzhen, China
[3]Department of Child Psychiatry of Shenzhen Kangning Hospital, Shenzhen Mental Health Center, Shenzhen, China

Corresponding Author:
Xingda Qu, PhD
Institute of Human Factors and Ergonomics
College of Mechatronics and Control Engineering
Shenzhen University
3688 Nanhai Avenue
Shenzhen, 518000
China
Phone: 86 86965716
Email: quxd@szu.edu.cn

Abstract

Background: Previous studies have shown promising results in identifying individuals with autism spectrum disorder (ASD) by applying machine learning (ML) to eye-tracking data collected while participants viewed varying images (ie, pictures, videos, and web pages). Although gaze behavior is known to differ between face-to-face interaction and image-viewing tasks, no study has investigated whether eye-tracking data from face-to-face conversations can also accurately identify individuals with ASD.

Objective: The objective of this study was to examine whether eye-tracking data from face-to-face conversations could classify children with ASD and typical development (TD). We further investigated whether combining features on visual fixation and length of conversation would achieve better classification performance.

Methods: Eye tracking was performed on children with ASD and TD while they were engaged in face-to-face conversations (including 4 conversational sessions) with an interviewer. By implementing forward feature selection, four ML classifiers were used to determine the maximum classification accuracy and the corresponding features: support vector machine (SVM), linear discriminant analysis, decision tree, and random forest.

Results: A maximum classification accuracy of 92.31% was achieved with the SVM classifier by combining features on both visual fixation and session length. The classification accuracy of combined features was higher than that obtained using visual fixation features (maximum classification accuracy 84.62%) or session length (maximum classification accuracy 84.62%) alone.

Conclusions: Eye-tracking data from face-to-face conversations could accurately classify children with ASD and TD, suggesting that ASD might be objectively screened in everyday social interactions. However, these results will need to be validated with a larger sample of individuals with ASD (varying in severity and balanced sex ratio) using data collected from different modalities (eg, eye tracking, kinematic, electroencephalogram, and neuroimaging). In addition, individuals with other clinical conditions (eg, developmental delay and attention deficit hyperactivity disorder) should be included in similar ML studies for detecting ASD.

(J Med Internet Res 2021;23(8):e29328) doi:10.2196/29328

KEYWORDS

autism spectrum disorder; eye tracking; face-to-face interaction; machine learning; visual fixation

Introduction

Autism spectrum disorder (ASD) is a complex neurodevelopmental condition characterized by social communication deficits along with restricted and repetitive behavior [1]. Owing to a lack of objective biomarkers, the current diagnosis of ASD heavily depends on behavioral evaluation, which involves substantive subjective procedures that can be negatively impacted by various factors such as caregivers' reporting bias and clinicians' insufficient capability in differentiating ASD [2,3]. In addition, the current diagnostic procedure is highly labor- and time-demanding due to the shortage in clinical specialists and requirement of lengthy examinations. A delayed diagnosis directly leads to postponed interventions, which subsequently impacts the prognosis of the affected children [4]. Therefore, seeking quantifiable and objective biomarkers of ASD, which could potentially make the diagnostic procedure more efficient and effective, has become a critical issue.

With respect to seeking objective biomarkers for ASD, recent studies reflect increasing interest in applying machine learning (ML) algorithms to examine whether features extracted from neuroimaging [5,6], electroencephalogram (EEG) [7], eye tracking [8,9], and kinematic data [10-12] could be used to identify ASD. The underlying justification for applying ML is based on the advantages of these approaches in identifying patterns that are not readily recognized by human eyes. Indeed, an ML approach demonstrated promising results in detecting ASD with objectively measured features. For example, Crippa et al [11] showed that seven kinematic features computed from a goal-directed motor task could accurately classify children with and without ASD (accuracy 96.7%). By implementing an imitation task, Li et al [13] reported a maximum classification accuracy of 86.7% using an ML approach.

Recently, a few studies have revealed that eye-tracking data could be used to identify ASD by implementing ML algorithms [8,9,14-19]. For example, Wan et al [9] recruited children within the ages of 4-6 years with ASD and typical development (TD) to watch a 10-second video displaying a woman speaking. ML features were extracted from eye-tracking measures in seven areas of interest (AOIs). Their results demonstrated that fixation time at the mouth and body AOIs could discriminate these two groups of participants with a classification accuracy of 85.1%. In contrast to Wan et al [9], who used a predefined AOI approach, Liu et al [8] used the *K*-means algorithm to extract features from the fixation data, which reached a maximum classification accuracy of 88.51%. Further, a few studies demonstrated that eye-tracking data obtained from web-searching tasks could be used to detect ASD [14-16]. Instead of computing features from eye-tracking data, Eraslan et al [15] performed a scan-path trend analysis to identify representative eye movement sequences for both individuals with ASD and TD. A classification was made based on the similarity of the individual's visual scan path to the representative sequences. This approach was able to classify individuals with ASD and TD with above-chance accuracy.

The eye-tracking data used in these prior studies were primarily obtained by having participants watch images (ie, videos, pictures, web pages) [8,9,14]. However, in reality, human gaze behavior is highly context-sensitive. Existing findings show that experimental settings and cognitive load are critical factors that could influence how people visually attend [20,21]. In contrast to image-watching tasks, face-to-face interaction is a social task that is much more perceptually and cognitively difficult [22]. Other studies have shown that the presence of the social partner elicits a different pattern of both neural response and gaze behavior [23,24]. In this vein, findings obtained from image-viewing tasks could not be directly generalized to the scenario of natural social interaction. Accordingly, there is a need to investigate whether eye-tracking data from live social interaction could be used to identify ASD.

The major novelty of this study is that we investigated the feasibility of using eye-tracking data from face-to-face conversations to classify children with ASD and TD. This research question is of practical significance since face-to-face interaction is omnipresent in everyday life. With the development of eye-tracking technology that enables the detection of natural social gaze behavior, ASD might be initially screened in daily life without needing to undergo lengthy and sophisticated procedures in clinical settings. In addition, apart from visual fixation measures, we included the length of conversation as an input feature to investigate whether combining features from these two modalities would increase the classification performance. The majority of prior eye-tracking ML research focused on using gaze data to identify ASD. To the best of our knowledge, only two recent studies combined eye tracking and EEG or kinematic data, showing that combined features yielded better classification performance than using features from a single modality [19,25]. With the development of objective assessment, it is proposed that future detection of ASD might be realized by integrating data from different modalities. Our research therefore contributes to the existing literature by investigating whether combining data from visual fixation and length of conversation could improve the performance of ML models.

Methods

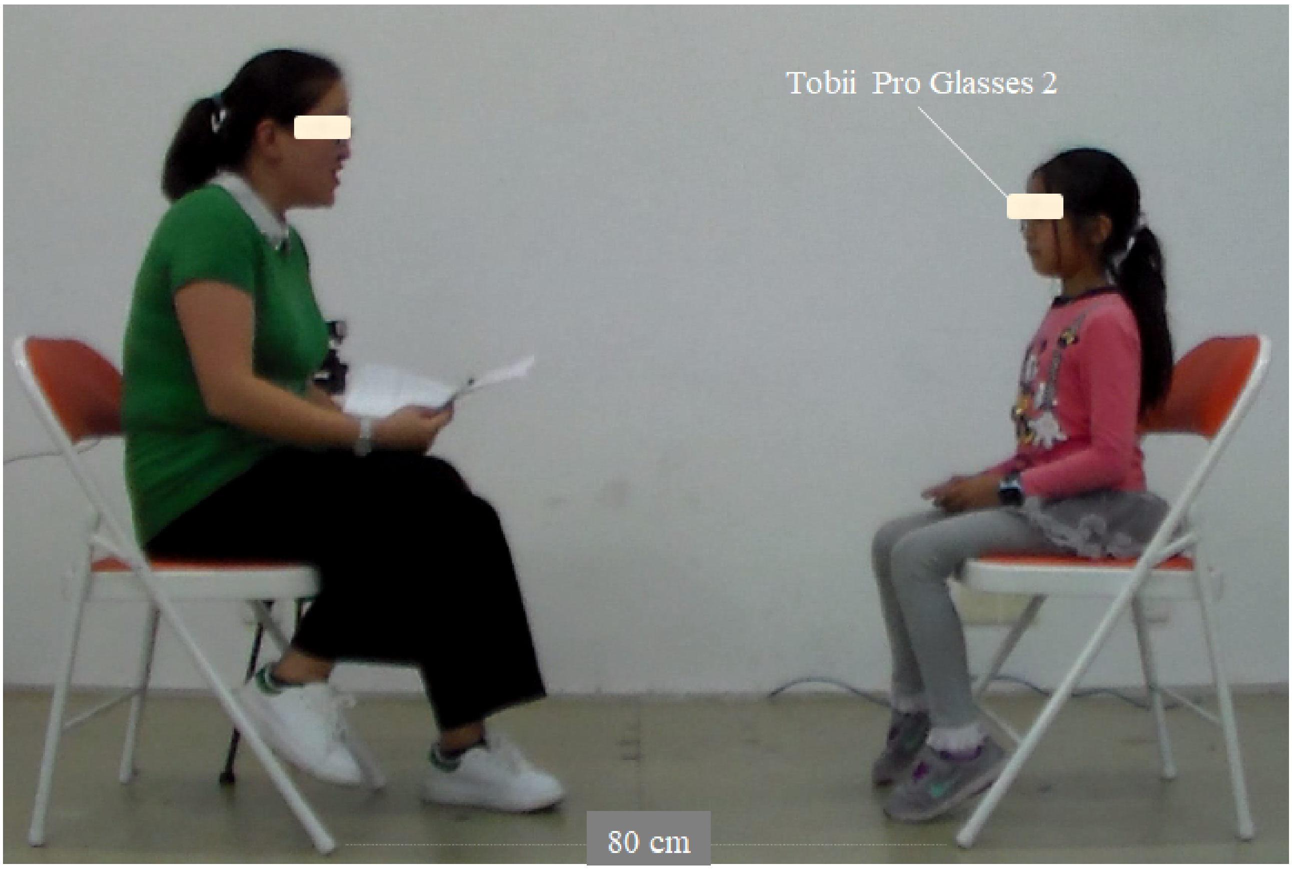
Participants

Data used in this study were obtained from a research project aiming at identifying behavioral markers of ASD. Twenty children with ASD and 23 children with TD were enrolled in the study. Children with ASD were recruited from the Child Psychiatry Department of Shenzhen Kangning Hospital. Owing to limited access to instruments such as the Autism Diagnostic Observation Schedule or the Autism Diagnostic Interview-Revised, ASD was primarily diagnosed by a licensed psychiatrist with no less than 5 years of clinical experience following the Diagnostic and Statistical Manual of Mental Disorders-IV criteria. In addition, the ASD diagnosis was further evaluated by a senior psychiatrist. A consultation with at least two additional senior psychiatrists would be arranged if there was disagreement among the specialists. All of these procedures ensured the correctness of the ASD diagnosis for the children enrolled in our study. Additional inclusion criteria were as follows: (1) aged between 6 and 13 years; (2) at least average nonverbal intelligence (IQ level was initially screened by the psychiatrist, and measured with the Raven advanced progressive matrices [26]); and (3) absence of other clinical conditions, including attention deficit hyperactivity disorder (ADHD) and schizophrenia. The TD group included healthy children without any mental or physical disorders and no diagnosis of ASD/ADHD in first-degree relatives, who were recruited from local schools. The experimental protocol followed the principles of the Declaration of Helsinki and the ethical guidelines of Shenzhen University. Written informed consent was provided by the participants' caregivers.

Data Collection

Participants were asked to engage in a structured face-to-face conversation with a 33-year-old female interviewer who was blinded to the participant's group membership. The interviewer was required to behave consistently across all interviews with all participants. Participants were required to wear a head-mounted eye tracker (Tobii Pro Glasses 2; sampling rate: 50 Hz; Tobii Technology, Stockholm, Sweden) during the conversation, and they were seated 80 cm away from the interviewer's chair (Figure 1). The conversation was videotaped with two still cameras. One camera (Samsung HMX-F90, sampling frequency 25 Hz) recorded both the interviewer and interviewee by placing each person equally on the left and right side of the recording view. The other camera (Logitech C270, sampling frequency 30Hz) was positioned beside the interviewer to capture the participant's behavior from the front view.

Figure 1. Experimental setup.



Participants were not informed of the function of the eye tracker, and they were asked to avoid moving the glasses or to make any intense head movements during the conversation. A postexperiment interview confirmed that none of the participants was aware that their gaze behavior had been recorded. In addition, once the eye tracker was moved by the participant (particularly those with ASD), an accuracy test was performed at the end of the conversation to ensure the accuracy of the eye-tracking data recording. Verifications showed that Tobii Pro Glasses 2 was reliably accurate even if the glasses were moved by participants during the conversation.

The structured conversation consisted of four chronologically arranged sessions: general questions in the first session, hobby sharing in the second session, yes-no questions in the third session, and question raising in the fourth session. The first session allowed both the interviewer and the child to become familiarized with each other. The second session served the purpose of examining the participants' behavior when speaking about their hobbies, which might induce different gaze behavior from that induced when discussing more generic topics [20]. The third session was designed to investigate the extent to which participants used head nodding or shaking to answer yes-no questions. The behavior of taking initiatives to raise questions was examined in the fourth session. Refer to Textbox 1 for further details of the questions used in each session.

Textbox 1. Details of the four sessions of the structured conversation.

Session 1: General questions

1. What is your name?
2. How is your name written?
3. What is the name of your school and what grade are you in?
4. Who is your best friend? What is your favorite thing to do together?
5. Could you please share with me the most interesting thing that happened last week? Let me know the time, place, people, and the whole process of the event.
6. What is the plan for your summer vacation?

Session 2: Hobby sharing

1. What is your favorite thing to do? And can you tell me why you like doing it?

Session 3: Yes-no questions

1. Do you like apples?
2. Do you like to go to the zoo?
3. Do you like to go to school?
4. Do you like reading?
5. Do you like painting?
6. Do you like watching cartoons?
7. Do you like sports?
8. Do you like watching movies?
9. Do you like traveling?
10. Do you like shopping?

Session 4: Question raising

1. Now that I have asked you many questions, do you have any questions for me?

Eye-Tracking Data Analysis

Data of four participants (one with ASD and three with TD) were discarded due to technical problems that occurred during the eye-tracking process. Hence, the final dataset consisted of 20 children with TD and 19 children with ASD. The participants' demographic information is presented in Table 1.

The eye-tracking data were analyzed with Tobii Pro Lab software, which enables processing visual fixation data on dynamic stimuli. Note that the interviewer was also a dynamic stimulus as she was interacting with the participants throughout the conversation.

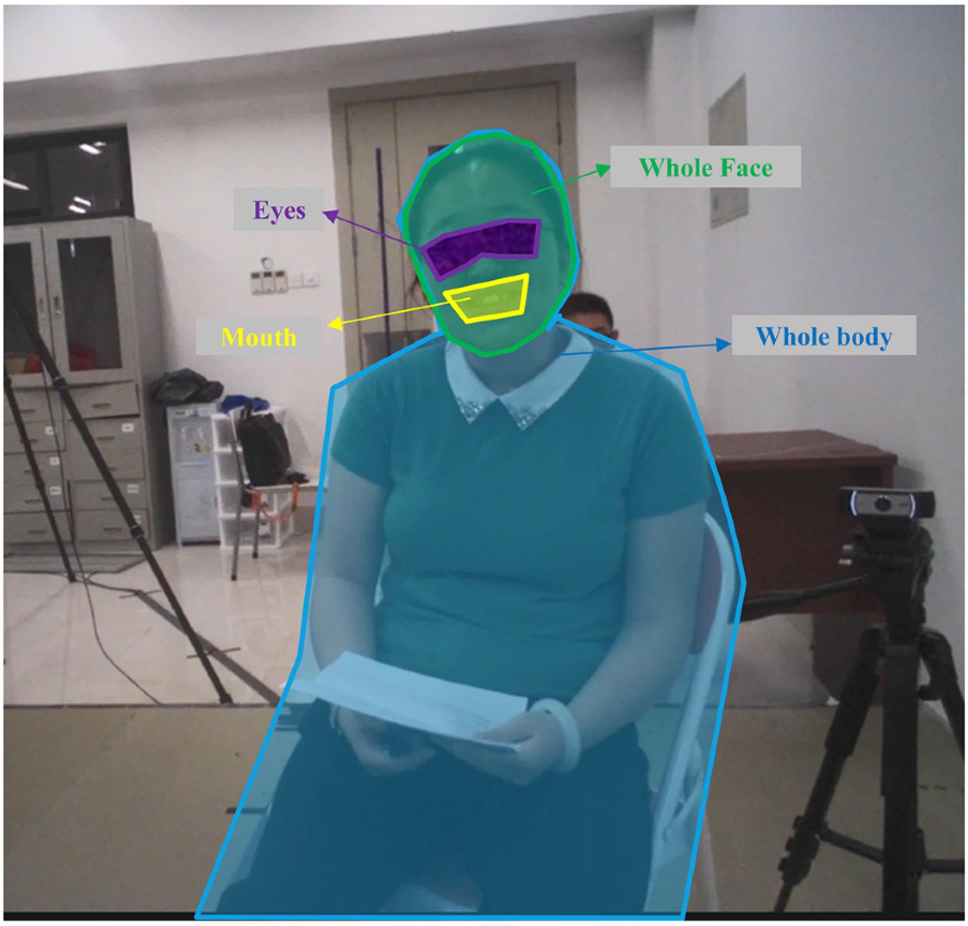
Features were extracted on visual fixation and session length from the eye-tracking data. For the visual fixation features, four AOIs were analyzed, including the eyes, mouth, whole face, and whole body (Figure 2). We computed the percentage of visual fixation time on each AOI as features. Therefore, 16 AOI-based features were acquired (4 sessions × 4 AOIs).

Table 1. Comparison of demographic information and the area of interest (AOI)-based features in the autism spectrum disorder (ASD) and typical development (TD) groups.

Characteristic	ASD	TD	Group comparison	*P* value
Demographic characteristics				
Sex ratio, M:F	17:2	17:3	χ^2_1=0.17	.68
Age (months), mean (SD)	99.6 (25.1)	108.8 (27.0)	t_{37}=1.09	.28
IQ, mean (SD)	100.8 (22.7)	116.1 (22.7)	t_{37}=2.45	.02
AOI features, mean (SD)[a]				
Mouth_Session 1	0.05 (0.06)	0.19 (0.13)	*U*=59.5	<.001
Eyes_Session 1	0.06 (0.06)	0.08 (0.09)	*U*=173.0	.63
Face_Session 1	0.21 (0.17)	0.41 (0.18)	*U*=70.5	.001
WholeBody_Session 1	0.33 (0.23)	0.55 (0.21)	*U*=85.5	.003
Mouth_Session 2	0.05 (0.09)	0.16 (0.13)	*U*=88.0	.004
Eyes_Session 2	0.04 (0.04)	0.06 (0.07)	*U*=143.0	.18
Face_Session 2	0.17 (0.16)	0.39 (0.20)	*U*=77.0	.001
WholeBody_Session 2	0.29 (0.26)	0.52 (0.25)	*U*=95.5	.008
Mouth_Session 3	0.12 (0.15)	0.21 (0.17)	*U*=131.0	.10
Eyes_Session 3	0.07 (0.06)	0.08 (0.10)	*U*=186.0	.91
Face_Session 3	0.33 (0.26)	0.49 (0.21)	*U*=120.5	.05
WholeBody_Session 3	0.46 (0.28)	0.06 (0.20)	*U*=134.5	.12
Mouth_Session 4	0.05 (0.06)	0.12 (0.12)	*U*=122.0	.05
Eyes_Session 4	0.06 (0.09)	0.08 (0.11)	*U*=183.5	.85
Face_Session 4	0.21 (0.20)	0.32 (0.18)	*U*=120.0	.05
WholeBody_Session 4	0.34 (0.25)	0.47 (0.22)	*U*=125.5	.07

[a] Due to a violation of the normality assumption, Mann-Whitney *U* tests were performed for group comparisons on AOI-based features.

Figure 2. Four areas of interest.



To obtain the percentage of visual fixation time on each AOI, the first step was to draw a snapshot image from the eye-tracking video for the purpose of defining AOIs. Once AOIs were defined, with the help of the real-world mapping algorithm, Tobii Pro Lab automatically mapped the gaze point in the video onto correct spots of the snapshot image. The correctness of the mapping process was confirmed by a human observer. Manual mapping was performed in the case that no fixation was automatically mapped onto the snapshot or if the fixation automatically assigned failed to match the correct spot. In this way, the accuracy of visual fixation was reliably ensured. Note that we used the velocity-threshold identification fixation filter to define fixation, which meant that a fixation was detected if the velocity of the eye movement was below 30 degrees per second for no less than 60 milliseconds. Finally, the percentage of visual fixation time on each AOI in a session was computed as the length of the fixation time on the AOI divided by the total duration of the particular session. Results regarding the group comparison on the AOI-based features in different sessions are presented in Table 1.

The length of each session varied across participants. Mann-Whitney *U* tests showed that the children with ASD had significantly longer conversations in the first session (*U*=48, *P*<.01), second session (*U*=103, *P*=.02), and fourth session (*U*=107, *P*=.02), but not in the third session (*U*=150, *P*=.26). In addition, the total length of all four sessions was significantly longer in the ASD group (*U*=39, *P*<.01). These results indicated that session length might serve as an effective feature to classify children with ASD and TD. Thus, the lengths of the four sessions and the total session length were used as five input features.

ML Procedure

Description of Dataset

Sixteen features on visual fixation (percentages of visual fixation time on four AOIs [mouth, eyes, face, and whole body] in four conversation sessions) and five features on session length were computed as features fed into the ML procedure. Therefore, the original dataset for the ML procedure was a 39 (participants)×21 (features) matrix. Three types of ML models were established, one with visual fixation features alone, one with session length features alone, and one with combined features on both modalities, to investigate whether combined features would yield better classification performance.

Classifiers

The classification task was performed by implementing four ML classifiers: support vector machine (SVM), linear discriminant analysis (LDA), decision tree (DT), and random forest (RF). The description of these classifiers is detailed below.

SVM is a supervised learning algorithm that has been previously implemented in classifying individuals with and without ASD [8,10]. The purpose of the SVM classifier is to create an optimal hyperplane in a multidimensional space with labeled training samples. Testing samples are classified based on the sign of the distance vector to the hyperplane, and the distance to the hyperplane determines the probability that they belong to the specific category.

The task of classifying children with ASD from those with TD is a binary classification problem. In this case, the LDA classifier works as a dimension reduction technique that projects all data points in the high-dimensional space onto a straight line (ie, one dimension) with training samples. Testing samples were classified in either group by the threshold value on the straight line.

The DT classifier is a tree-like flowchart. The nodes in the model represent tests on an attribute, the branches represent the outcomes of the tests, and the leaf nodes denote class labels. The DT classifier exhibits the advantage of strong interpretability, but it is prone to overfitting.

Instead of building a tree-like structure, the RF classifier is established by creating multiple simple trees with the training data. Test samples are categorized into a specific group based on the majority of votes from the trees.

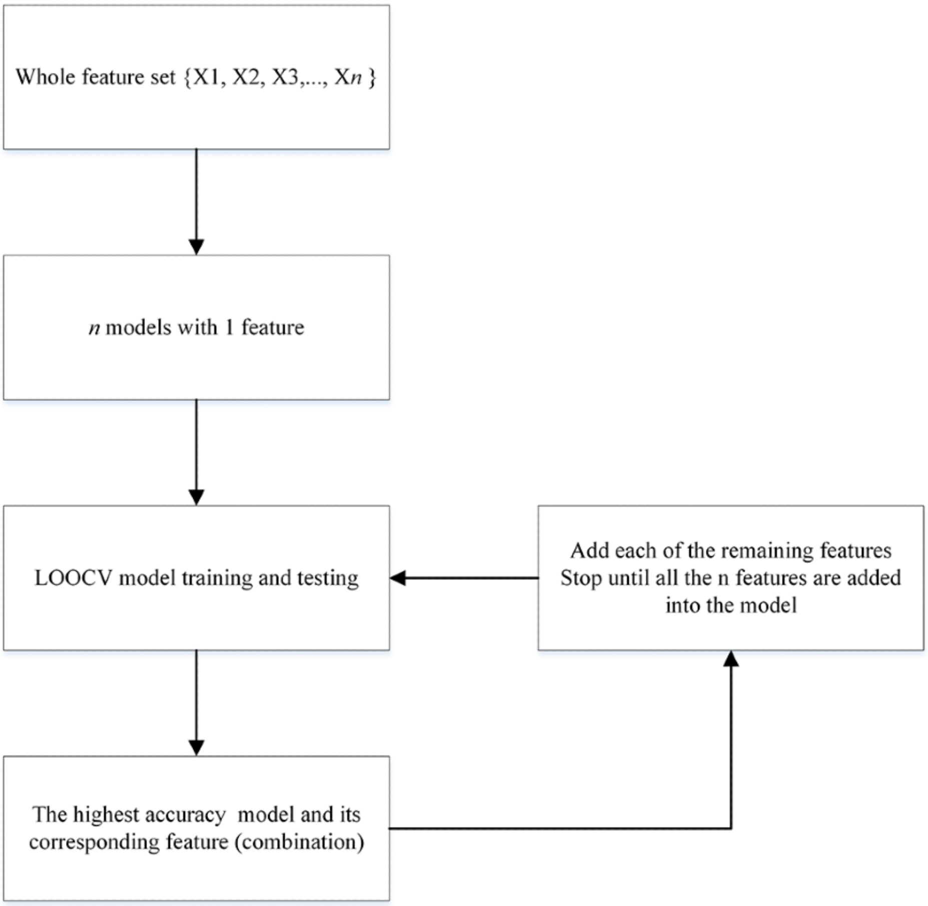
Feature Selection

Forward feature selection (FFS) was applied to select features for model training and testing. Specifically, FFS is an iterative process starting with the evaluation of each individual feature by examining their classification performance. The feature with the highest classification accuracy would be preserved and is then combined with each of the other features to form two-feature models whose classification performances are further evaluated. The two features with optimal classification accuracy are then retained and used to establish three-feature models by combining them with each of the remaining features. By repeating these procedures, the one-feature, two-feature, …, *n*-feature models with the highest classification accuracy would be obtained (*n* represents the total number of examined features intended to be fed into ML models). In this way, FFS helped to identify not only the model with the highest classification accuracy but also the corresponding feature or feature combination.

Classification

The entire ML procedure is schematically presented in Figure 3. To minimize the potential overfitting problem, we implemented leave-one-out cross-validation in ML model training and testing. Specifically, the test set contained only one participant sample and the remaining participant samples were used to train the ML model. This procedure was repeated until all participant samples were tested once. The accuracy, sensitivity, and specificity were computed to evaluate the classification of the ML models. Accuracy was defined as the percentage of participant samples that were correctly classified in both groups. Specificity and sensitivity corresponded to the model's capability of correctly detecting the TD and ASD samples respectively.

Figure 3. Flowchart of the machine learning procedure. LOOCV: leave-one-out cross-validation.

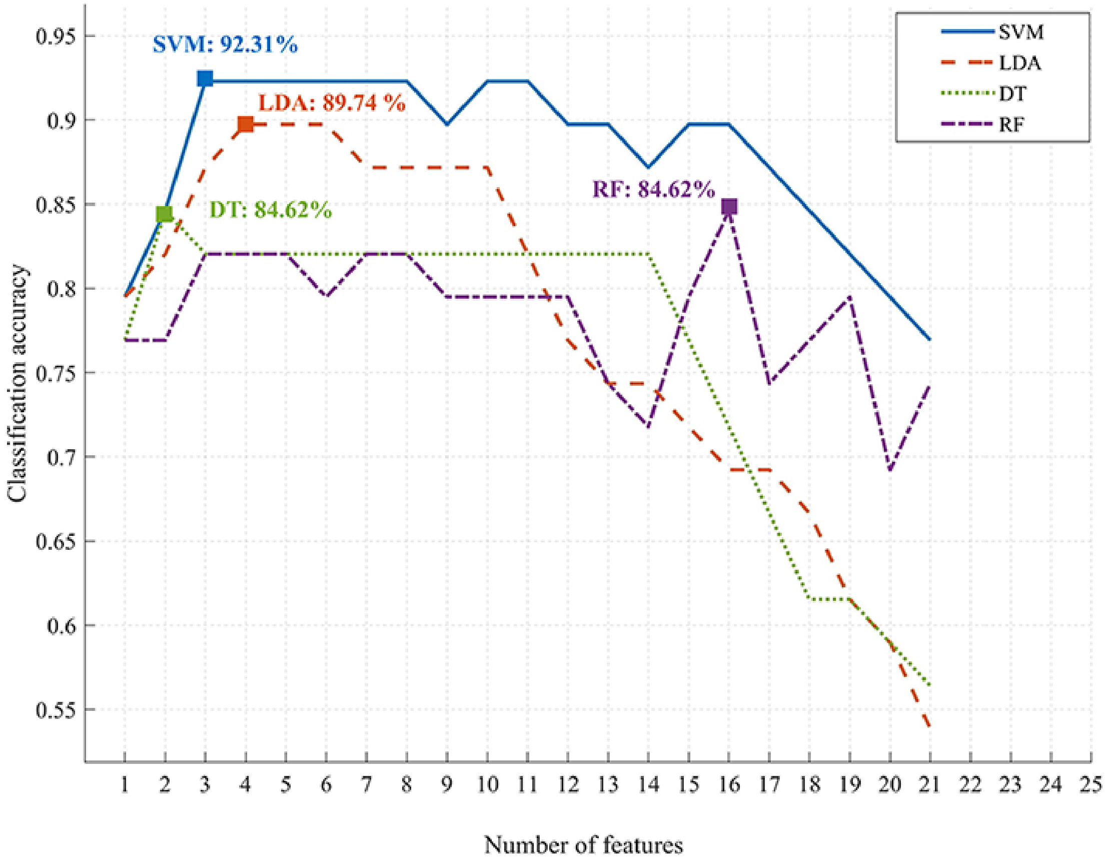

Whole feature set {X1, X2, X3,..., Xn }
n models with 1 feature
LOOCV model training and testing
Add each of the remaining features
Stop until all the n features are added
into the model
The highest accuracy model and its
corresponding feature (combination)

Results

Classification with Combined Features

The variation in classification accuracy according to the number of features used in the model is illustrated in Figure 4. All classifiers yielded a maximum classification accuracy above 84%. The SVM classifier achieved optimal classification accuracy of 92.31% with three features (specificity=100%, sensitivity=84.21%, area under the receiver operating characteristic curve [AUC]=0.92), followed by LDA with 89.74% accuracy using four features (specificity=90.00%, sensitivity=89.47%, AUC=0.92), DT with 84.62% accuracy using two features (specificity=80.00%, sensitivity=89.47%, AUC=0.86), and RF with 84.62% accuracy using 16 features (specificity=85.00%, sensitivity=84.21%, AUC=0.86).

Figure 4. Variation of the classification accuracy with the number of features. SVM: support vector machine; LDA: linear discriminant analysis; DT: decision tree; RF: random forest.



The classification performance of the SVM classifier was the highest among the four classifiers. The variation of the SVM classification performance according to the number of features is presented in Table 2. The classification accuracy reached 79.49% with only one feature: total session length. The optimal classification accuracy of 92.31% was achieved with a minimum of three features: total session length, mouth in the first session, and whole body in the third session.

Table 2. Variation of the support vector machine classification performance with different features.

Number of features	Added feature	Accuracy (%)	Sensitivity (%)	Specificity (%)
1	Total SL[a]	79.49	68.42	90.00
2	~[b] +Mouth_Session 1	84.62	78.95	90.00
3	~ +Wholebody_Session 3	92.31	84.21	100.00
4	~ +Face_Session 3	92.31	84.21	100.00
5	~ +Face_Session 2	92.31	89.47	95.00
6	~ +Eyes_Session 4	92.31	89.47	95.00
7	~ +Face_Session 1	92.31	89.47	95.00
8	~ +SL_Session 2	92.31	89.47	95.00
9	~ +Wholebody_Session 1	89.74	89.47	90.00
10	~ +Face_Session 4	92.31	89.47	95.00
11	~ +Mouth_Session 2	92.31	89.47	95.00
12	~ +Eyes_Session 1	89.74	84.21	95.00
13	~ +Eyes_Session 2	89.74	84.21	95.00
14	~ +Mouth_Session 3	87.18	84.21	90.00
15	~ +SL_Session 3	89.74	84.21	95.00
16	~ +Wholebody_Session 4	89.74	84.21	95.00
17	~ +Mouth_Session 4	87.18	84.21	90.00
18	~ +Eyes_Session 3	84.62	78.95	90.00
19	~ +SL_Session 1	82.05	78.95	85.00
20	~ +SL_Session 4	79.49	78.95	80.00
21	~ +Wholebody_Session 2	76.92	73.68	80.00

[a] SL: session length.

[b] In forward feature selection, ~ represents all features in the previous iteration; for example, ~ represents all 6 previously selected features in the 7th iteration.

The confusion matrix of this three-feature model that achieved the highest accuracy is presented in Table 3, which shows that the model correctly classified children in the TD group with 100% accuracy, but it mislabeled three children with ASD as having TD. Error analysis examining the mislabeled samples showed that these participants performed equally well as the children with TD (Figure 5). For example, the total session length of mislabeled sample 1 was shorter than that of 75% of the children in the TD group, and the visual fixation time on the mouth AOI in the first session was higher than that of half of the children in the TD group. Consistent with a previous study [27], these results support the significant heterogeneity among individuals with ASD.

Table 3. Confusion matrix of the support vector machine classifier with the highest accuracy.[a]

Actual class	Predicted class	
	TD[b]	ASD[c]
TD	TN[d]=20	FP[e]=0
ASD	FN[f]=3	TP[g]=16

[a] Accuracy=TP+TN/TP+FP+FN+TN; sensitivity=TP/TP+FN; specificity=TN/FP+TN.

[b] TD: typical development.

[c] ASD: autism spectrum disorder.

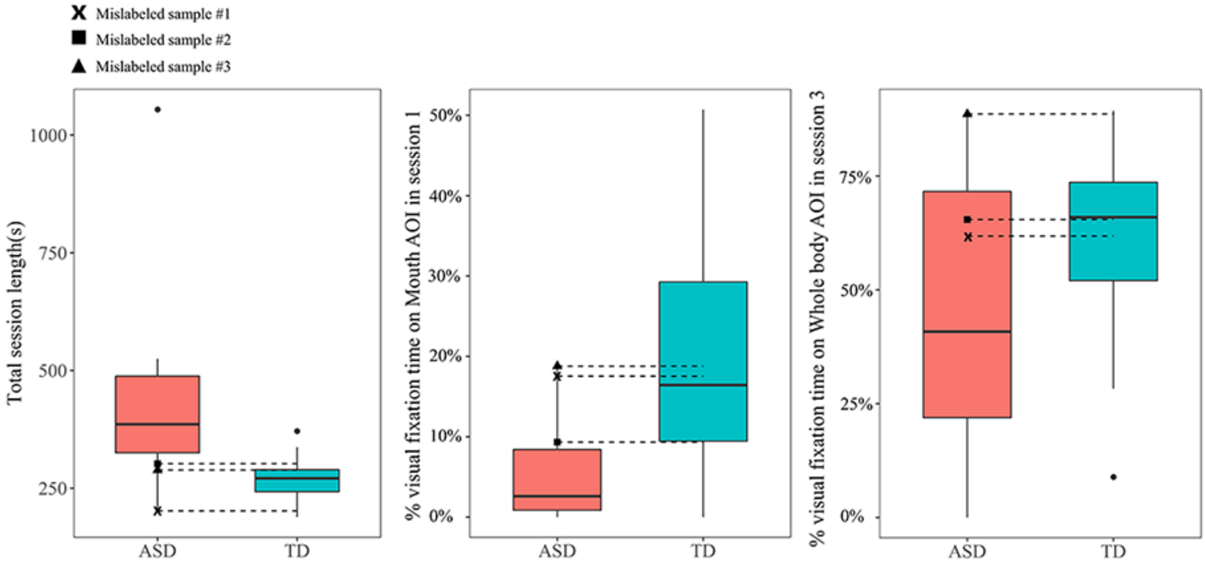
[d] TN: true negative.

[e] FP: false positive.

[f] FN: false negative.

[g] TP: true positive.

Figure 5. Boxplots of three features that achieved the highest classification accuracy in the support vector machine classifier along with the three mislabeled samples. ASD: autism spectrum disorder; TD: typical development.



Classification Using Only Visual Fixation Features

Following the same procedure but feeding only AOI-based features into the ML classifiers achieved a maximum classification accuracy of 84.62% by the LDA classifier (specificity=80.00%, sensitivity=89.47%, AUC=0.86) with three features (mouth in session 1, face in session 2, and mouth in session 3), and by the DT classifier (specificity=80.00%, sensitivity=89.47%, AUC=0.86) with two features (face in session 2 and eyes in session 3).

Classification Using Only Session Length Features

When using only session length features to perform the classification task, the maximum classification accuracy of 84.62% was achieved by the SVM classifier (specificity=90.00%, sensitivity=78.95%, AUC=0.87) with four features (session length in sessions 1, 3, and 4, and total session length).

Discussion

Principal Findings

In this study, we extracted features on visual fixation and session length from eye-tracking data collected during face-to-face conversations and investigated their capacity for classifying children with ASD and TD. The maximum classification accuracy of 92.31% was achieved by combining features on both visual fixation and session length with the SVM classifier. The classification accuracy was higher than that obtained using visual fixation features (highest accuracy: 84.62%) or session length features (highest accuracy: 84.62%) alone. Since 19 children with ASD and 20 children with TD were enrolled in this study, there was a slight class imbalance. Majority class prediction is typically used as a baseline for imbalanced classification. In the context of this study, majority class prediction requires every participant sample to be predicted as “TD”. Thus, the classification accuracy of majority class prediction would be 51.3% (ie, 20/39), which is greatly lower than the optimal classification accuracy of our results. This suggests that our results could not be explained by majority class prediction.

The highest classification accuracy was achieved with three features: total session length, percentage of visual fixation time on the mouth AOI in the first session, and percentage of visual fixation time on the whole body AOI in the third session. As shown in Table 2, the total session length was an effective feature for discriminating ASD from TD with an accuracy of 79.49% alone. In our study, participants were engaged in a structured conversation, in which they had to interact with the interviewer by answering the same number of questions. Longer conversation might be explained by the social deficits in children with ASD. Specifically, it was assumed that children with ASD might have experienced greater difficulty in understanding the social information (eg, motivation, mental state, and emotion) conveyed by the interviewer [28,29]. Interestingly, various studies demonstrated that the social deficits are more pronounced when dealing with naturalistic social stimuli [29,30]. Thus, it took the children with ASD longer to finish the same number of questions. However, further exploration is needed to confirm whether the length of conversation could be attributed to the poor social understanding capacity.

Notably, fixation measures on the mouth and whole body AOIs played important roles in the SVM classifier that produced the highest classification accuracy. The mouth AOI emerged as a prominent feature in this study, possibly owing to the fact that participants were engaged in a conversational task. Previous studies showed that the mouth is an important body feature that affords the looking-toward behavior in conversations [22,31,32]. Our result of selecting the mouth AOI as an important feature was consistent with the findings of Wan et al [9], in which participants watched a video of a model speaking. With respect to the whole body AOI, abundant research has shown that individuals with ASD pay less attention to socially relevant stimuli [33,34]. The interviewer in this study could be viewed as the most relevant social stimulus, as participants needed to utilize information of the interviewer (eg, emotions, gestures, body movements) to converse with her. Looking away from the interviewer would induce the missing of important social information, which may further undermine the ability of the participants with ASD to interact with the interviewer during the conversation.

Apart from the fact that we used data from face-to-face interaction as opposed to data obtained from image-viewing tasks used in previous related studies, our study is different from other eye-tracking ML studies in two main aspects. First, this study recruited children aged between 6 and 13 years, whereas Wan et al [9] studied younger children (4-6 years old) and other studies [14-16,19] tested the adult population. Age is of profound significance in this context, since early identification and intervention may tremendously improve the prognosis of individuals with ASD [4]. A recent meta-analysis reported that the mean age at diagnosis of ASD was 60.48 months and was 43.18 months when only incorporating children aged ≤10 years [35]. This suggests that future ML studies should focus on examining younger children to facilitate the detection of ASD at an early stage. Second, the ASD severity level was not specifically measured in our study, which was accounted for in a previous study [27]. The children with ASD included in this study could be viewed as representing individuals with minor severity. It is recommended that individuals with ASD with different degrees of severity be included in future studies to improve the generalizability of the ML model. Except for these two differences, it is notable that our study and most others only classified individuals with ASD and TD [8,9,14-16]. Therefore, it remains unclear whether eye-tracking data could effectively detect ASD from other clinical phenotypes (eg, developmental delay and ADHD). More scientific endeavor is certainly required before a practical ML model that could detect ASD from different conditions is established.

Limitations

To ensure that the participants would be able to converse with the interviewer, we recruited children within the age range of 6-13 years with at least average intellectual ability. Participants with severe symptoms of autism were not included. In addition, only four girls were enrolled in our study. Prior studies reported that males with ASD differ from females with ASD in many respects, including behavioral presentation, cognitive domains, and emotions [36,37]. Therefore, this study should only be considered as proof-of-concept research, which explored the feasibility of using eye-tracking data from face-to-face conversations to classify children with ASD and TD. Future studies might consider recruiting participants with various presentations (eg, different degrees of severity and balanced sex ratio) to ensure the generalizability of the ML model.

This study utilized a head-mounted eye tracker to record the gaze behavior, which might affect the social behavior of children with ASD to a larger extent. In general, individuals with ASD are more sensitive to wearing devices and eye-tracking techniques usually require extensive calibration [38,39]. These issues considerably raise the difficulty of implementing eye-tracking techniques on children with ASD, particularly on the younger population. To address these problems, a recent study used a webcam to record eye movement and developed a computer vision–based algorithm to detect gaze behavior. The results showed that the accuracy of the algorithm was comparable to that of manual coding when evaluating particular gaze behaviors [39]. It is proposed that more contactless and calibration-free techniques should be developed to record the gaze behavior in individuals with ASD.

Our study only computed the percentage of visual fixation time on different AOIs as measures of gaze behavior. In fact, a variety of other features could be obtained from the gaze behavior, including the number of fixations, entropy, and number of revisits [16,40]. Additionally, features extracted from oculomotor behavior are also recommended since atypical oculomotor performance has been extensively reported in individuals with ASD [41,42]. Future ML studies are encouraged to generate as many features as possible so as to allow for specification of the globally optimal set of features for ASD identification.

Using eye-tracking data from face-to-face interaction was a major novelty of this study. However, human interaction may introduce a variety of subjective factors that are difficult to control but might influence the gaze behavior of participants. For example, the interviewer might unconsciously behave differently with the children with ASD from the TD group, even if she was required to maintain a similar manner of behavior when interacting with participants in both groups. To examine whether the interviewer behaved consistently with both groups of participants, the overall amount of movement she made during the conversation was estimated using image differencing techniques applied to the video recordings [43,44]. Statistical analysis of these data showed that the amount of the interviewer's movement was not significantly different when interacting with these two groups of participants (t_{214}=1.76, *P*=.29). However, it is acknowledged that a similar amount of body movement does not necessarily mean that the interviewer's behavior was completely identical for all participants. This is an inevitable problem faced by all studies investigating natural social interaction since no human being can be expected to behave exactly the same way when interacting with different people. In summary, future studies attempting to apply eye tracking to live social interactions need to cautiously control for factors (eg, context, task, and the interactant's behavior) that might be introduced through human interaction.

Conclusion

Our study extracted features from eye-tracking data during face-to-face conversations to investigate their capacity of detecting children with ASD. With a relatively small sample, our results showed that combining features on visual fixation and session length could accurately classify children with ASD and those with TD. It is proposed that future eye-tracking ML studies could use features from gaze-based measures [8,9], visual scanning path [15], and oculomotor performance [41,42] to detect ASD. Finally, we recommend that a larger and younger participant sample should be tested with the ML approach by combining features obtained from different modalities (eye tracking, neuroimaging, EEG, and kinematic) to evaluate how these objectively measured features could contribute to the early screening of ASD.

Acknowledgments

This study was financially supported by the SZU funding project (number 860-000002110259), Science and Technology Innovation Committee of Shenzhen (number JCYJ20190808115205498), Key Medical Discipline of GuangMing Shenzhen (number 12 Epidemiology), Sanming Project of Medicine in Shenzhen (number SZSM201612079), Key Realm R&D Program of Guangdong Province (number 2019B030335001), Shenzhen Key Medical Discipline Construction Fund (number SZXK042), and Shenzhen Double Chain Grant (number [2018]256).

Authors' Contributions

ZZ, XZ, XQ, and JL designed the experiment. ZZ, HT, and XH performed the data analyses. ZZ, HT, and XQ wrote the manuscript.

Conflicts of Interest

None declared.

References

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (5th ed.). Arlington, VA: American Psychiatric Publishing; 2013.
2. Möricke E, Buitelaar JK, Rommelse NNJ. Do we need multiple informants when assessing autistic traits? The degree of report bias on offspring, self, and spouse ratings. J Autism Dev Disord 2016 Jan;46(1):164-175 [FREE Full text] [doi: 10.1007/s10803-015-2562-y] [Medline: 26334871]
3. Tebartz van Elst L, Pick M, Biscaldi M, Fangmeier T, Riedel A. High-functioning autism spectrum disorder as a basic disorder in adult psychiatry and psychotherapy: psychopathological presentation, clinical relevance and therapeutic concepts. Eur Arch Psychiatry Clin Neurosci 2013 Nov;263(Suppl 2):S189-S196. [doi: 10.1007/s00406-013-0459-3] [Medline: 24105433]
4. Dawson G, Jones EJ, Merkle K, Venema K, Lowy R, Faja S, et al. Early behavioral intervention is associated with normalized brain activity in young children with autism. J Am Acad Child Adolesc Psychiatry 2012 Nov;51(11):1150-1159 [FREE Full text] [doi: 10.1016/j.jaac.2012.08.018] [Medline: 23101741]
5. Plitt M, Barnes KA, Martin A. Functional connectivity classification of autism identifies highly predictive brain features but falls short of biomarker standards. Neuroimage Clin 2015;7:359-366 [FREE Full text] [doi: 10.1016/j.nicl.2014.12.013] [Medline: 25685703]
6. Yamagata B, Itahashi T, Fujino J, Ohta H, Nakamura M, Kato N, et al. Machine learning approach to identify a resting-state functional connectivity pattern serving as an endophenotype of autism spectrum disorder. Brain Imaging Behav 2019 Dec;13(6):1689-1698. [doi: 10.1007/s11682-018-9973-2] [Medline: 30280304]
7. Ibrahim S, Djemal R, Alsuwailem A. Electroencephalography (EEG) signal processing for epilepsy and autism spectrum disorder diagnosis. Biocyber Biomed Eng 2018;38(1):16-26. [doi: 10.1016/j.bbe.2017.08.006]
8. Liu W, Li M, Yi L. Identifying children with autism spectrum disorder based on their face processing abnormality: A machine learning framework. Autism Res 2016 Aug;9(8):888-898. [doi: 10.1002/aur.1615] [Medline: 27037971]
9. Wan G, Kong X, Sun B, Yu S, Tu Y, Park J, et al. Applying eye tracking to identify autism spectrum disorder in children. J Autism Dev Disord 2019 Jan;49(1):209-215. [doi: 10.1007/s10803-018-3690-y] [Medline: 30097760]
10. Zhao Z, Zhang X, Li W, Hu X, Qu X, Cao X, et al. Applying machine learning to identify autism with restricted kinematic features. IEEE Access 2019;7:157614-157622. [doi: 10.1109/access.2019.2950030]
11. Crippa A, Salvatore C, Perego P, Forti S, Nobile M, Molteni M, et al. Use of machine learning to identify children with autism and their motor abnormalities. J Autism Dev Disord 2015 Jul;45(7):2146-2156. [doi: 10.1007/s10803-015-2379-8] [Medline: 25652603]
12. Anzulewicz A, Sobota K, Delafield-Butt JT. Toward the autism motor signature: gesture patterns during smart tablet gameplay identify children with autism. Sci Rep 2016 Aug 24;6(1):31107. [doi: 10.1038/srep31107] [Medline: 27553971]
13. Li B, Sharma A, Meng J, Purushwalkam S, Gowen E. Applying machine learning to identify autistic adults using imitation: An exploratory study. PLoS One 2017;12(8):e0182652 [FREE Full text] [doi: 10.1371/journal.pone.0182652] [Medline: 28813454]
14. Yaneva V, Ha LA, Eraslan S, Yesilada Y, Mitkov R. Detecting high-functioning autism in adults using eye tracking and machine learning. IEEE Trans Neural Syst Rehabil Eng 2020 Jun;28(6):1254-1261. [doi: 10.1109/tnsre.2020.2991675]
15. Eraslan S, Yesilada Y, Yaneva V, Harper S. Autism detection based on eye movement sequences on the web: a scanpath trend analysis approach. 2020 Presented at: The 17th International Web for All Conference; April 2020; Taipei, Taiwan p. 1-10. [doi: 10.1145/3371300.3383340]
16. Yaneva V, Ha LA, Eraslan S, Yesilada Y, Mitkov R. Detecting autism based on eye-tracking data from web searching tasks. 2018 Presented at: 15th International Web for All Conference; April 2018; Lyon, France p. 1-10. [doi: 10.1145/3192714.3192819]

17. Oliveira JS, Franco FO, Revers MC, Silva AF, Portolese J, Brentani H, et al. Computer-aided autism diagnosis based on visual attention models using eye tracking. Sci Rep 2021 May 12;11(1):10131. [doi: 10.1038/s41598-021-89023-8] [Medline: 33980874]
18. Akter T, Ali MH, Khan MI. Machine learning model to predict autism investigating eye-tracking dataset. 2021 Presented at: 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST); Dhaka, Bangladesh; February 2021. [doi: 10.1109/icrest51555.2021.9331152]
19. Vabalas A, Gowen E, Poliakoff E, Casson AJ. Applying machine learning to kinematic and eye movement features of a movement imitation task to predict autism diagnosis. Sci Rep 2020 May 20;10(1):8346. [doi: 10.1038/s41598-020-65384-4] [Medline: 32433501]
20. Nadig A, Lee I, Singh L, Bosshart K, Ozonoff S. How does the topic of conversation affect verbal exchange and eye gaze? A comparison between typical development and high-functioning autism. Neuropsychologia 2010 Jul;48(9):2730-2739. [doi: 10.1016/j.neuropsychologia.2010.05.020] [Medline: 20493890]
21. Doherty-Sneddon G, Bruce V, Bonner L, Longbotham S, Doyle C. Development of gaze aversion as disengagement from visual information. Dev Psychol 2002;38(3):438-445. [doi: 10.1037/0012-1649.38.3.438]
22. Klin A, Jones W, Schultz R, Volkmar F, Cohen D. Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. Arch Gen Psychiatry 2002 Sep;59(9):809-816. [doi: 10.1001/archpsyc.59.9.809] [Medline: 12215080]
23. Freeth M, Foulsham T, Kingstone A. What affects social attention? Social presence, eye contact and autistic traits. PLoS One 2013;8(1):e53286 [FREE Full text] [doi: 10.1371/journal.pone.0053286] [Medline: 23326407]
24. Pönkänen LM, Alhoniemi A, Leppänen JM, Hietanen JK. Does it make a difference if I have an eye contact with you or with your picture? An ERP study. Soc Cogn Affect Neurosci 2011 Sep;6(4):486-494 [FREE Full text] [doi: 10.1093/scan/nsq068] [Medline: 20650942]
25. Kang J, Han X, Song J, Niu Z, Li X. The identification of children with autism spectrum disorder by SVM approach on EEG and eye-tracking data. Comput Biol Med 2020 May;120:103722. [doi: 10.1016/j.compbiomed.2020.103722] [Medline: 32250854]
26. Raven J. The Raven's progressive matrices: change and stability over culture and time. Cogn Psychol 2000 Aug;41(1):1-48. [doi: 10.1006/cogp.1999.0735] [Medline: 10945921]
27. Jacob S, Wolff JJ, Steinbach MS, Doyle CB, Kumar V, Elison JT. Neurodevelopmental heterogeneity and computational approaches for understanding autism. Transl Psychiatry 2019 Feb 04;9(1):63. [doi: 10.1038/s41398-019-0390-0] [Medline: 30718453]
28. Byrge L, Dubois J, Tyszka JM, Adolphs R, Kennedy DP. Idiosyncratic brain activation patterns are associated with poor social comprehension in autism. J Neurosci 2015 Apr 08;35(14):5837-5850 [FREE Full text] [doi: 10.1523/JNEUROSCI.5182-14.2015] [Medline: 25855192]
29. Golan O, Baron-Cohen S, Golan Y. The 'Reading the Mind in Films' Task [child version]: complex emotion and mental state recognition in children with and without autism spectrum conditions. J Autism Dev Disord 2008 Sep;38(8):1534-1541. [doi: 10.1007/s10803-007-0533-7] [Medline: 18311514]
30. Heavey L, Phillips W, Baron-Cohen S, Rutter M. The Awkward Moments Test: a naturalistic measure of social understanding in autism. J Autism Dev Disord 2000 Jun;30(3):225-236. [doi: 10.1023/a:1005544518785] [Medline: 11055458]
31. Gibson J. The ecological approach to visual perception. Boston: Houghton Mifflin; 1979.
32. Jones RM, Southerland A, Hamo A, Carberry C, Bridges C, Nay S, et al. Increased eye contact during conversation compared to play in children with autism. J Autism Dev Disord 2017 Mar;47(3):607-614. [doi: 10.1007/s10803-016-2981-4] [Medline: 27987063]
33. Frazier TW, Strauss M, Klingemier EW, Zetzer EE, Hardan AY, Eng C, et al. A meta-analysis of gaze differences to social and nonsocial information between individuals with and without autism. J Am Acad Child Adolesc Psychiatry 2017 Jul;56(7):546-555 [FREE Full text] [doi: 10.1016/j.jaac.2017.05.005] [Medline: 28647006]
34. Chita-Tegmark M. Social attention in ASD: A review and meta-analysis of eye-tracking studies. Res Dev Disabil 2016 Jan;48:79-93. [doi: 10.1016/j.ridd.2015.10.011] [Medline: 26547134]
35. van 't Hof M, Tisseur C, van Berckelear-Onnes I, van Nieuwenhuyzen A, Daniels AM, Deen M, et al. Age at autism spectrum disorder diagnosis: A systematic review and meta-analysis from 2012 to 2019. Autism 2021 May;25(4):862-873 [FREE Full text] [doi: 10.1177/1362361320971107] [Medline: 33213190]
36. Mandy W, Chilvers R, Chowdhury U, Salter G, Seigal A, Skuse D. Sex differences in autism spectrum disorder: evidence from a large sample of children and adolescents. J Autism Dev Disord 2012 Jul;42(7):1304-1313. [doi: 10.1007/s10803-011-1356-0] [Medline: 21947663]
37. Bölte S, Duketis E, Poustka F, Holtmann M. Sex differences in cognitive domains and their clinical correlates in higher-functioning autism spectrum disorders. Autism 2011 Jul;15(4):497-511. [doi: 10.1177/1362361310391116] [Medline: 21454389]
38. Dufour M, Lanovaz MJ. Increasing compliance with wearing a medical device in children with autism. J Appl Behav Anal 2020 Apr;53(2):1089-1096. [doi: 10.1002/jaba.628] [Medline: 32319096]

39. Vargas-Cuentas NI, Roman-Gonzalez A, Gilman RH, Barrientos F, Ting J, Hidalgo D, et al. Developing an eye-tracking algorithm as a potential tool for early diagnosis of autism spectrum disorder in children. PLoS One 2017;12(11):e0188826 [FREE Full text] [doi: 10.1371/journal.pone.0188826] [Medline: 29190703]
40. Wang Q, Hoi SP, Wang Y, Song C, Li T, Lam CM, et al. Out of mind, out of sight? Investigating abnormal face scanning in autism spectrum disorder using gaze-contingent paradigm. Dev Sci 2020 Jan;23(1):e12856. [doi: 10.1111/desc.12856] [Medline: 31081980]
41. Nowinski CV, Minshew NJ, Luna B, Takarae Y, Sweeney JA. Oculomotor studies of cerebellar function in autism. Psychiatry Res 2005 Nov 15;137(1-2):11-19. [doi: 10.1016/j.psychres.2005.07.005] [Medline: 16214219]
42. Miller M, Chukoskie L, Zinni M, Townsend J, Trauner D. Dyspraxia, motor function and visual-motor integration in autism. Behav Brain Res 2014 Aug 01;269:95-102 [FREE Full text] [doi: 10.1016/j.bbr.2014.04.011] [Medline: 24742861]
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. Alviar C, Dale R, Dewitt A, Kello C. Multimodal coordination of sound and movement in music and speech. Discourse Process 2020 Jun 18;57(8):682-702. [doi: 10.1080/0163853x.2020.1768500]

Abbreviations

ADHD: attention deficit hyperactivity disorder
AOI: area of interest
ASD: autism spectrum disorder
AUC: area under the receiver operating characteristic curve
DT: decision tree
EEG: electroencephalogram
FFS: forward feature selection
LDA: linear discriminant analysis
ML: machine learning
RF: random forest
SVM: support vector machine
TD: typical development

Edited by R Kukafka; submitted 02.04.21; peer-reviewed by V Yaneva, Y Man; comments to author 05.06.21; revised version received 30.06.21; accepted 05.07.21; published 26.08.21.

Please cite as:
Zhao Z, Tang H, Zhang X, Qu X, Hu X, Lu J
Classification of Children With Autism and Typical Development Using Eye-Tracking Data From Face-to-Face Conversations: Machine Learning Model Development and Performance Evaluation
J Med Internet Res 2021;23(8):e29328
URL: https://www.jmir.org/2021/8/e29328
doi:10.2196/29328
PMID:34435957

©Zhong Zhao, Haiming Tang, Xiaobin Zhang, Xingda Qu, Xinyao Hu, Jianping Lu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Anticipated Benefits and Concerns of Sharing Hospital Outpatient Visit Notes With Patients (Open Notes) in Dutch Hospitals: Mixed Methods Study

Sharon L Janssen[1,2], MSc; Nynke Venema-Taat[2], BSc; Stephanie Medlock[1], DVM, PhD

[1]Department of Medical Informatics, Amsterdam Public Health Research Institute, Amsterdam UMC, University of Amsterdam, Amsterdam, Netherlands
[2]EvA Servicecentrum, Amsterdam UMC, University of Amsterdam, Amsterdam, Netherlands

Corresponding Author:
Stephanie Medlock, DVM, PhD
Department of Medical Informatics, Amsterdam Public Health Research Institute
Amsterdam UMC
University of Amsterdam
Meibergdreef 9
Amsterdam, 1105 AZ
Netherlands
Phone: 31 205664543
Email: s.k.medlock@amsterdamumc.nl

Abstract

Background: The past few years have seen an increase in interest in sharing visit notes with patients. Sharing visit notes with patients is also known as “open notes.” Shared notes are seen as beneficial for patient empowerment and communication, but concerns have also been raised about potential negative effects. Understanding barriers is essential to successful organizational change, but most published studies on the topic come from countries where shared notes are incentivized or legally required.

Objective: We aim to gather opinions about sharing outpatient clinic visit notes from patients and hospital physicians in the Netherlands, where there is currently no policy or incentive plan for shared visit notes.

Methods: This multimethodological study was conducted in an academic and a nonacademic hospital in the Netherlands. We conducted a survey of patients and doctors in March-April 2019. In addition to the survey, we conducted think-aloud interviews to gather more insight into the reasons behind participants’ answers. We surveyed 350 physicians and 99 patients, and think-aloud interviews were conducted with an additional 13 physicians and 6 patients.

Results: Most patients (81/98, 77%) were interested in viewing their visit notes, whereas most physicians (262/345, 75.9%) were opposed to allowing patients to view their visit notes. Most patients (54/90, 60%) expected the notes to be written in layman’s terms, but most physicians (193/321, 60.1%) did not want to change their writing style to make it more understandable for patients. Doctors raised concerns that reading the note would make patients feel confused and anxious, that the patient would not understand the note, and that shared notes would result in more documentation time or losing a way to communicate with colleagues. Interviews also revealed concerns about documenting sensitive topics such as suspected abuse and unlikely but worrisome differential diagnoses. Physicians also raised concerns that documenting worrisome thoughts elsewhere in the record would result in fragmentation of the patient record. Patients were uncertain if they would understand the notes (46/90, 51%) and, in interviews, raised questions about security and privacy. Physicians did anticipate some benefits, such as the patients remembering the visit better, shared decision-making, and keeping patients informed, but 24% (84/350) indicated that they saw no benefit. Patients anticipated that they would remember the visit better, feel more in control, and better understand their health.

Conclusions: Dutch patients are interested in shared visit notes, but physicians have many concerns that should be addressed if shared notes are pursued. Physicians’ concerns should be addressed before shared notes are implemented. In hospitals where shared notes are implemented, the effects should be monitored (objectively, if possible) to determine whether the concerns raised by our participants have actualized into problems and whether the anticipated benefits are being realized.

(J Med Internet Res 2021;23(8):e27764) doi:10.2196/27764

KEYWORDS

patient portal; access to information; barriers and facilitators; survey; qualitative; open notes; mixed methods; electronic health record

Introduction

Background

One of the defining changes in patient care in the past few decades is the rise of the concept of patient empowerment. The World Health Organization defines empowerment as “a process through which people gain greater control over decisions and actions affecting their health” [1]. The operationalization of empowerment takes many forms, including interventions to improve patients’ knowledge and health literacy, applications and devices for better self-management, and advocating shared decision-making. An important aspect of empowerment is greater transparency in the health care process. This viewpoint is reflected in statements such as the National Academy of Medicine 2001 recommendation that “patients should have unfettered access to their own medical information” [2], and in Dutch law, which requires that patients be given a copy of their record upon request, and that electronic access should now be offered [3]. These directives have been interpreted in various ways. Many hospitals and clinics worldwide now offer patients web-based access to information such as current medications and laboratory results. However, a more controversial question is whether access to the medical record should include access to doctors’ free-text visit notes. Free-text visit notes are notes that a clinician writes about a patient’s visit in the patient record, as opposed to structured information such as lab values. Access to visit notes (and not just structured data such as lab results) is viewed as part of a movement toward greater transparency in health care [4].

The content of visit notes varies, but typically, visit notes contain the doctor’s observations, assessment (including differential diagnoses), and plan for treatment or further diagnostics. In 2010, Beth Israel Deaconess Medical Center captained a study of “open notes,” wherein 114 primary care providers experimented with shared notes by giving patients full access to their visit notes [4]. The results of this experiment were positive, with patients reporting positive effects and experiences, and clinicians reporting minimal disruption to their work [5], despite initial concerns about negative effects on documentation and taking too much time from clinicians and staff [4]. Since then, a number of health care institutions worldwide have adopted shared notes. These studies also report benefits from shared notes, such as patients feeling better prepared for clinic visits [6,7], feeling more in control of their health [7,8], and feeling that they better recall the doctor’s instructions [7,8]. However, some concerns emerged as well: patients reported difficulty understanding the notes [7], patients were offended by some content in the notes [6,9], or clinicians reported omitting information from the notes out of concern that it might offend the patient [8].

Most studies are based in the United States [6-10] and a few from Sweden [11]. In 2015, the United States entered stage 3 of Meaningful Use, which is a federal program that encourages the use of health information technology. Stage 3 requires the adoption of shared notes to receive financial incentives. Similarly, access to clinic notes is considered a right in Sweden, and most regions in Sweden have implemented shared notes in some form [12]. These incentives may lead to a more positive view of shared notes. The concerns and benefits may be different in other countries due to differences in culture, health care systems, or other differences. In the Netherlands, a patient must be given a copy of his or her full record upon request, but there is no requirement for the visit notes to be made available on the web. Hospitals in the Netherlands must have a patient portal available, so the visit note could be made available via the portal. A pilot study of shared notes has been announced, but the results have not yet been published [13].

Objectives

Little is known about Dutch patients’ interest in shared notes, the benefits anticipated by physicians or patients, or their concerns. It is important to understand these barriers before attempting an implementation and to ensure that patients and physicians have realistic expectations of the benefits. Adding the Dutch perspective can help broaden the understanding currently reflected in the literature. Therefore, our aim is to assess the attitudes of patients and physicians regarding shared outpatient visit notes in a Dutch hospital setting and elucidate their anticipated benefits and concerns by means of a survey and interviews.

Methods

Overview

We chose a multimethodological (mixed methods) approach, using a short quantitative survey to gather opinions from a large number of patients and practitioners and think-aloud interviews to confirm participants’ interpretation of the survey questions and to understand the nuances behind the responses. As sharing notes with inpatients during hospitalization poses additional technical and practical challenges, we focused on shared visit notes in the outpatient setting. Technical challenges include providing equipment to view notes; practical challenges include determining the appropriate delay before releasing notes, providing bedside technical support services, ensuring privacy, and other challenges. The content of inpatient notes also differs from that of outpatient clinic notes. Both benefits and concerns are expected to differ between inpatient and outpatient settings.

Survey

Development

The surveys were developed using published surveys on the topic [14-16] as a starting point. One survey contained questions about potential benefits and concerns [14], whereas the other two investigated only benefits [15,16]. Relevant questions from these studies were identified by one researcher (SLJ) and confirmed by a second researcher (SM). These questions were adapted, and new questions were added by the researchers based on the aims of our study (concerns, benefits, and attitudes toward shared visit notes). The survey was iteratively discussed and revised by the research team until all team members were satisfied with the questions.

Pilot Testing

The patient survey was piloted with health communication experts. The physician survey was piloted with doctors who were familiar with the procedures in the participating hospitals and the electronic patient record but were not eligible as subjects for the survey (ie, not currently practicing in the participating hospitals).

The feedback from the pilot tests were used to make the final survey for the patients (29 questions over 6 pages) and physicians (23 questions over 4 pages).

Participant Selection for Surveys

Physicians

The physician survey was sent to doctors from both an academic hospital (the same hospital where the patient survey was conducted) and a nonacademic hospital in the same region. Both hospitals use the same electronic health record system. An email was sent to all heads of all outpatient departments at the academic hospital and to a contact person at the nonacademic hospital who was asked to distribute it to the heads of departments there. The heads of the departments were asked to forward the email to all physicians working in their outpatient department. The email contained a short description of the study and the link to the survey. In addition, before sending the email, the study was introduced at a meeting of the heads of the outpatient departments at the academic hospital. The physicians' survey was deployed on the web, using a custom form written in the PHP (PHP: Hypertext Preprocessor) programming language, and made available for 5 weeks (March 25, 2019, to May 1, 2019). No reminders were sent. Survey responses from the think-aloud interviews (described below) were added to the survey results by hand.

Patients

The patient survey was distributed in the outpatient clinic of a large academic hospital in the Netherlands. Adult, Dutch-speaking patients who attended the outpatient clinic were invited in person by a researcher (SLJ) to participate in the survey. Arrangements for the researcher to attend the outpatient clinic were made with the team leaders of the outpatient clinic, who are responsible for personnel in the outpatient clinic. From March 27, 2019, till April 16, 2019, the researcher went to various outpatient departments to hand out the surveys to patients. The researcher invited consecutive patients arriving in the waiting area on the days that she attended. We selected departments to include both older and younger patients, patients with chronic and acute disease, and varying seriousness of their diagnoses. The researcher (SLJ) approached the patient in the waiting rooms and introduced herself, the study, and the duration of the survey (5-10 minutes). If the patient agreed to participate, the researcher handed out the survey on paper and left the patient to fill in the survey. The same method was used to recruit patients for interviews; the first patients from each waiting area who were approached for the survey were asked to complete the survey with the researcher present in a think-aloud interview. Data on patients who declined to participate was not collected. Data from the paper surveys were transcribed to a spreadsheet by one researcher (SLJ).

Aggregation and Coding of Data

The data analysis consisted of simple counts and percentages. Patients and physicians were allowed to skip questions; therefore, we analyzed each question with n equal to the number of responses to that question. Responses to open questions were coded by one researcher (SLJ) using open coding (manifest content analysis), and the codes were discussed with a second researcher (SM).

The introductory text of both the patient and physician surveys informed participants about the study and that all data would be stored and processed anonymously. Both surveys were voluntary, and no incentives were offered to the patients or physicians.

Think-Aloud Interviews

Overview

Following the methods of Westerman et al [17], we asked a subsample of physicians and patients to fill in the survey, and “think aloud” about their reasoning while filling in the answers. The researcher asked prompting questions if the respondent did not explain their answer out loud. All think-aloud interviews were performed by the first author (SLJ), a master's student in medical informatics. This researcher's experience with surveys and qualitative research included courses and an internship during her bachelor's and master's degrees. The patients and physicians were informed of the name and background of the first author and were informed of the reason for this study and asked to participate anonymously. The researcher had no relationship with the patients or physicians before the study period. All interviews were audio-recorded and transcribed by the first author.

Participant Selection for Think-Aloud Interviews

Physicians

In the email used to invite physicians to participate in the survey at the academic hospital, doctors were also invited to contact the researchers to participate in an interview. Interviews were continued until saturation was reached in the responses, and all physicians who responded were interviewed.

Patients

As part of the process of distributing the surveys, if an extra exam room was available in the outpatient clinic, the researcher invited patients to take the survey in the room and “think aloud” while completing it. We used a purposive sampling method: 1 or 2 patients were invited to be interviewed in each department visited while distributing the surveys until saturation was reached in the interview results.

Aggregation and Coding of Data

The transcribed recordings were coded by a single researcher (SLJ) using thematic analysis. A predetermined starting set of codes was used based on the constructs underlying the survey questions. Open coding was used to classify items that did not fit in the predetermined set. Two coded interviews were checked by a second researcher (SM).

Ethics

The Medical Ethics Committee issued a waiver for this study, indicating that it does not fall under the Human Research Law of the Netherlands and that no further ethical approval is needed. Informed consent was obtained from all subjects.

Results

The survey instruments are given in Multimedia Appendix 1 [14-16] and annotated with the references used in developing the surveys. The original surveys were in Dutch; they were translated to English by a native English speaker (SM).

Survey

Physician Survey

Participants

A total of 350 physicians completed all (321/350, 91.7%) or a part (29/350, 8.2%) of the survey. An additional 15 empty responses (where the survey was viewed but no questions were answered) were excluded. The demographics of the participants are presented in Table 1. For physicians who were interviewed who had not completed the survey at the time of the interview, their responses to the survey during the interview were counted as part of the survey responses (8/350, 2.2%). Of the two participating hospitals, 72.8% (255/350) were from the academic hospital, and 17.4% (61/350) were from the nonacademic hospital (the remaining 34/350, 9.7% of respondents skipped this item).

Table 1. Demographic characteristics of the physicians (N=350)[a].

Characteristics	Responses, n (%)
Gender (n=314[b])	
Female	174 (55.4)
Male	138 (43.9)
Other	2 (0.6)
Age (years; n=314)	
18-28	6 (1.9)
29-39	117 (37.2)
40-49	100 (31.8)
50-59	72 (22.9)
60-69	19 (6.1)
Location (n=316)	
Academic hospital	255 (80.6)
Nonacademic hospital	61 (19.3)
Department (n=320)	
Other (free text)	91 (28.4)
Gynecology, obstetrics, and gender	27 (8.4)
Hematology	25 (7.8)
Lung department	24 (7.5)
Internal medicine	23 (7.1)
Cardiology	21 (6.5)
Surgery	18 (5.6)
Departments with <5% of participants each (12 departments)	67 (20.9)

[a] N indicates the total number of participants who filled in any demographic information.

[b] The n for each question indicate the number of participants who answered that specific question.

Respondents who selected “other department” could fill in a free textbox; the most common department given in this group (11/320, 3.4%) was the pediatric medicine department.

Opinions on Sharing Notes: Physician Survey

Most physicians in this survey would prefer not to share the visit notes with patients (282/345, 81.7%). When asked what information is in the visit notes, physicians indicated that their visit notes contain the anamnesis (343/350, 98%), treatment plan (339/350, 96.8%), diagnosis (328/350, 93.7%), physical examination (325/350, 92.8%), interpretation and/or summary (321/350, 91.7%), differential diagnosis (315/350, 90.0%), laboratory results (314/350, 89.7%), additional examinations (315/350, 90%), and medical history (283/350, 80.8%). The subjects that physicians were most concerned about sharing with patients were the differential diagnosis (196/350, 56%), interpretation and/or summary (162/350, 46.2%), and anamnesis (110/350, 31.4%).

Reasons why the physicians would not like to share the notes or a part of the note with the patients are given in Table 2. Participants were able to select any number of responses.

Table 2. Reasons why physicians do not want to share (part of) the visit note (N=350)[a].

Reason	Responses[b], n (%)
It will make the patient confused and anxious	236 (67.4)
The patient will not understand the content of the note	221 (63.1)
Not all of the information is relevant for the patient	205 (58.5)
I have not (yet) spoken with the patient about all the information in the note	201 (57.4)
Some information in the note is only relevant for colleagues	165 (47.1)
I expect many more questions from the patient	154 (44)
It will generate more work	122 (34.8)
The visit note is my personal note	113 (32.2)

[a] The total “N” indicates the total number of physicians who responded to this question.

[b] Number of participants choosing this response.

We also allowed respondents to fill in free-text reasons why they did not want to share (part of) the visit notes, which were added by 20% (70/350) of respondents. Analysis of these free-text comments underscored concerns about confusing and worrying the patient, especially by reading the differential diagnosis. Doctors pointed out that they often need to consider the possibility of an unlikely but worrisome diagnosis such as cancer or amyotrophic lateral sclerosis (Lou Gehrig disease) and may want to watch for signs of it on diagnostic tests but do not want to discuss it with the patient unless there is a substantial chance that the patient actually has this disease. Doctors also remarked that the notes are a place to record their thoughts so they can pick up their train of thought later on and that these thoughts might not be complete or may later be proven wrong. This is especially problematic when the doctor suspects a sensitive problem, such as abuse or sexually transmitted diseases. The record also functions as a tool to facilitate discussion of these matters with colleagues. Doctors also expressed concern about family members reading the file or that information would be left out or displaced to other parts of the record, thus compromising the quality of care.

Physicians were largely unwilling to write their notes with less jargon and abbreviations (193/321, 60.1%) to make it more understandable for patients. A smaller group of physicians were willing to partly change their writing style (85/321, 26.4%), and the rest were willing to change their writing style (43/321, 13.3%).

Benefits and Concerns: Physician Survey

In addition to asking physicians about their reasons for sharing or not sharing their visit notes, we also asked physicians about the anticipated benefits of sharing notes with patients and about their concerns. Although there is some overlap with reasons for not sharing notes, we asked about them separately as a doctor may have a concern but not consider it a reason to avoid sharing notes. Participants were able to select any number of responses. The results are presented in Tables 3 and 4.

Table 3. Anticipated benefits according to physicians (N=322)[a].

Potential benefit	Responses[b], n (%)
The patient can better remember and understand what was said during the visit	144 (44.7)
It facilitates shared decision making	98 (30.4)
The patient is better informed about their illness and health	96 (29.8)
The patient can make corrections or additions	82 (25.4)
The patient can put the already-accessible results[c] in context	47 (14.5)
The family of the patient will be better informed	46 (14.2)
It facilitates a better doctor-patient relationship	37 (11.4)
Patient compliance will increase	23 (7.1)

[a] The total "N" indicates the total number of physicians who responded to this question.

[b] Number of participants choosing this response.

[c] Patients already have access to laboratory results via the patient portal.

Table 4. Anticipated concerns according to physicians (N=322[a]).

Potential concern	Responses[b], n (%)
The patient will be confused because they won't understand the content	276 (85.7)
The patient will be worried (eg, by the differential diagnosis)	276 (85.7)
I will need to answer more questions via the portal	239 (74.2)
The patient will ask for corrections and/or additions	215 (66.7)
I will have difficulty communicating with my colleagues because some of the information would be perceived differently by the patient	203 (63)
It will generate more work	200 (62.1)
I will need to spend more time on documentation	200 (62.1)
The family of the patient will interfere	130 (40.3)
I will need to explain more to the patient during the consult	117 (36.3)

[a] The total "N" indicates the total number of physicians who responded to this question.

[b] Number of participants choosing this response.

We also offered a free textbox where physicians could fill in other benefits or concerns, which was filled by 28.8% (93/322) of the participants. Analysis of the free-text responses showed that 23.9% (77/322) of physicians said that they believed that there are no benefits for the patients or physicians. Other comments (16/322, 4.9%) underscored the benefits of increased retention of information, self-efficacy, and the importance of transparency, as well as the concern that patients will not understand what is written.

Patient Survey

Participants

A total of 99 patients participated in the survey. The demographics of the respondents are presented in Table 5.

Table 5. Demographic characteristics of patients (N=99)[a].

Characteristics	Responses, n (%)
Gender (n=94)[b]	
Female	68 (72)
Male	26 (28)
Age (years; n=94)	
<18	1 (1)
18-28	17 (18)
29-39	15 (16)
40-49	10 (11)
50-59	22 (23)
60-69	20 (21)
70-79	9 (10)
Highest education level (n=83)	
Primary school	1 (1)
High school	39 (46)
University	43 (52)
Self-reported health (n=88)	
Very poor	2 (2)
Poor	20 (23)
Fine	51 (58)
Very well	13 (15)
Excellent	2 (2)

[a] The total "N" indicates the number of participants who filled in any demographic information.

[b] The "n" value for each question indicates the number of participants who answered that specific question.

Opinions on Sharing Notes: Patient Survey

In contrast to the physicians, most patients found it important (50/90, 56%) or very important (19/90, 21%) to read the visit notes in the patient portal. Patients were most interested in seeing their laboratory results (78/89, 88%), a summary of the visit (71/89, 80%), and the diagnosis or differential diagnosis (60/89, 67%). Other parts that interested patients were the treatment plan (45/89, 51%), medication (41/89, 46%), and physical examination (34/89, 38%).

Patients agreed (27/90, 30%) or strongly agreed (28/90, 31%) that they expected notes to be written in understandable language or layman's terms if the doctors know that they will be reading the notes. Patients indicated that if they did not understand the notes, they would ask the doctor at the next visit (39/90, 43%), discuss with family or friends (16/90, 18%), call the department (12/90, 13%), or send a message to the doctor (10/90, 11%). We also asked the patients what they would do if they were to see information in the note that they disagreed with or did not expect. Patients indicated that they would ask the doctor about it during the next visit (62/90, 69%), call the department (38/90, 42%), or send a message to the doctor (22/90, 24%).

Patients were also asked about what they would do with the information in the notes, the benefits they anticipate from open notes, and their concerns (Table 6).

Table 6. Responses to statements on anticipated responses to notes, possible benefits, and concerns about shared notes.

Statement	Strongly disagree, n (%)	Disagree, n (%)	Neutral, n (%)	Agree, n (%)	Strongly agree, n (%)
Discuss the notes with my doctor or another doctor (n=89[a])	2 (2)	12 (14)	25 (28)	39 (44)	11 (12)
Discuss the notes with family and friends (n=90)	8 (9)	16 (18)	27 (30)	32 (35)	7 (8)
I would better remember what was discussed at the visit (n=90)	3 (3)	2 (2)	14 (16)	50 (56)	21 (23)
I would feel more in control of the care process (n=90)	4 (4)	5 (6)	26 (29)	40 (44)	15 (17)
I would better understand my illness and health (n=90)	5 (5)	23 (26)	30 (33)	42 (47)	12 (13)
The notes will be more confusing than useful (n=90)	8 (9)	36 (40)	32 (35)	8 (9)	6 (7)
I would worry more about my health (n=89)	9 (10)	42 (47)	29 (33)	6 (7)	3 (3)

[a] The “n” value for each statement indicates the number of participants who responded to that specific question.

Think-Aloud Interviews

Physician Interviews

Participants

In total, 13 physicians were interviewed. The duration of the physician interviews was 20-25 minutes. The interview participants were 38% (5/13) male and 62% (8/13) female and worked in 10 different departments.

Reasons for Not Wanting to Share Visit Notes: Physician Interviews

Physicians commented on their reasons for not wanting to share the visit notes with patients (Table 7). The most common reason was fear that the information would be confusing or misleading for the patient:

> *As a dermatologist we know 3000 skin diseases. So for each spot, an experienced doctor could think of 10 to 20 diagnoses. This would be very confusing for the patient.*

Table 7. Coded reasons for not sharing visit notes (n=13).

Reason	Participants, n (%)[a]
Will be confusing or patient would not understand	7 (54)
Not relevant for patient	6 (46)
Insecurities with differential diagnosis	5 (38)
Sensitive data, nuances, or psychological and social information	4 (31)
Considerations	4 (31)
No possibility of a note with colleagues	4 (31)
Did not discuss everything	3 (23)
Discuss when I am 100% sure	3 (23)
File for the doctor	3 (23)
No benefit for patient health care	3 (23)

[a] Number of participants who mentioned a reason classified under this code.

Physicians also mentioned the need to document sensitive data, including things such as the suspicion of domestic violence:

> *Now and then you are suspecting domestic violence or something else. Before you will mention it to the authorities, you need to be sure and gather some evidence.*

Physicians also mentioned the function of the patient record as a cognitive aid for the diagnostic process and the need to note clinical hunches that may not yet be confirmed:

> *I don’t have any secrets from my patients, never. So, from that point of view they are allowed to read the notes - but the point is that I need a space for considerations, worries, and fears. You need to write that somewhere.*

Benefits and Concerns: Physician Interviews

In addition to the concerns in the structured part of the survey, doctors mentioned two additional concerns. First, notes are sometimes entered by doctors in training, and they may not yet have the skills to communicate in a way that is appropriate for patients. Physicians also mentioned the need to document private conversations with minors and whether their parents would have access to the notes:

> *One of my patients just turned 18 and she has a worrying family situation. Her dad was very controlling and I think that they will make comments about my interpretation that it is not only her [that has a problem], because her diagnosis is supported by factors such as stressful family situation.*

Jargon: Physician Interviews

We asked the physicians if they were willing to write the notes in understandable language for the patients, and 6 of them said no. Four physicians mentioned that jargon is the most efficient way to work for them: "Every higher education, where intellectual effort is needed and everything that has to do with professionalism has jargon."

Differences in Visit Note Between Departments: Physician Interviews

In total, 3 physicians said that there is a difference between the notes of different departments. The example was given that notes about a fractured hip (orthopedics) are very different from the notes from hematology.

Other Ways to Achieve the Benefits: Physician Interviews

Half of the physicians wanted to have more time and communication with the patient; 3 physicians mentioned that they were writing information on paper during the visit. Physicians also mentioned sharing letters (sent to the general practitioner) or sharing the visit summary as better alternatives.

Patient Interviews

Participants

In total, 6 patients were interviewed. The average duration of the patient interviews was 10-15 minutes. Some patients asked if their partner could attend the interview, so in two interviews, nonparticipants were present. Participants were 50% (3/6) female and 50% (3/6) male, falling into five different age categories (ranging from 18-28 years to 70-79 years).

Reasons for Wanting Open Notes: Patient Interviews

All patients who participated in the interviews were interested in seeing the visit notes but named various reasons for being interested (Table 8). The most cited reason was curiosity; the second was feeling that the note was about them, and this itself is a reason to see it: "I would like seeing the notes because it is about me and I think it is very important that I know exactly what happened." Patients also mentioned that seeing the notes would help them remember what was discussed, especially after receiving bad news from the physician: "Nine out of ten times you will hear the half of what the doctor is saying." All 6 patients said that they were interested in seeing all parts of the note.

Table 8. Coded reasons for wanting to access visit notes (n=6).

Reason	Participant, n (%)[a]
Curious	4 (67)
It's about me	3 (50)
Medical background	3 (50)
Cannot remember everything	2 (33)
Bad news	2 (33)

[a] Number of participants who mentioned a reason classified under this code.

Use of Medical Jargon: Patient Interviews

Patients also had varied opinions about expecting a note to be in layman's terms (Table 9); 2 of the patients interviewed believed that the note should be in layman's terms:

> *If the note contains a lot of abbreviations than it makes no sense for the patient to read the note because I would not understand half of the note.*

However, 2 patients were neutral on the topic, and 2 disagreed:

> *No, I don't expect it because the doctors need to have the possibility to talk and speak in their jargon.*

Table 9. Comments on the use of medical jargon (n=6).

Comment	Participants, n (%)[a]
Doctors need jargon	3 (50)
I would not understand the notes	3 (50)
Already administrative burden	3 (50)
Works easier for doctors	2 (33)
Prefer it but do not expect it	1 (17)

[a] Number of participants who mentioned a reason classified under this code.

Benefits and Concerns: Patient Interviews

In total, 3 patients mentioned privacy as a potential concern with open notes, both in terms of internet security and who might be given access to the notes in addition to the patient themselves (eg, family members or home care workers).

Other Ways to Achieve the Benefits: Patient Interviews

Patients suggested that some of these benefits could be achieved by bringing someone with you to the visit and having good informational leaflets.

Discussion

Principal Findings

We designed and conducted surveys regarding opinions on shared notes from 350 physicians and 99 patients, and conducted interviews with 13 doctors and 6 patients. Of all participants, 81.7% (282/345) of doctors prefer not to share visit notes with patients. Physicians indicated that nearly all aspects of care appear in the note, and they were particularly concerned about patients reading the differential diagnosis, the interpretation or summary, and the anamnesis. The most common reasons were worries that reading the note would make patients confused and anxious, that the patient will not understand their notes, that the information is not relevant for the patient, and that the note may contain information that has not yet been discussed with the patient. Clinical notes are written using medical jargon, and most physicians (193/321, 60.1%) did not want to change their writing style to make it more understandable for patients. Physicians did anticipate some benefits, such as better patient recall of what was discussed, better shared decision-making, and keeping patients informed. However, 23.9% (77/322) indicated that they saw no benefit in allowing patients to access the visit notes. Physicians also had many concerns (with some overlap with their reasons for not wanting to share notes), including unnecessary confusion and worry for the patient and family, needing more time to answer patient questions and more time for documentation, and more difficulty communicating with colleagues via the notes. The interviews clarified that physicians were concerned about the need to document sensitive information, such as the suspicion of domestic violence, and the need to have a place to document conversations with minors. They also mentioned the function of the patient record as a cognitive aid to sort through unconfirmed thoughts. Physicians were also concerned about patients reading notes written by doctors in training, who might not write things in a way that is appropriate for the patients. In the patient survey, 77% (69/90) of patients found it important or very important to see their visit notes. Patients were most interested in viewing their laboratory results, visit summary, and diagnoses. Most patients (55/90, 60%) expected visit notes to be written in layman's terms. Most patients indicated that if they had questions, they would ask them at the next visit, although some (12/90, 13%) indicated that they would call the hospital to ask. A higher percentage indicated that they would call (38/90, 42%) or send a note (22/90, 24%) if they found information that they did not agree with or did not expect. Patients saw some potential benefit to reading their notes: they felt they would better remember what was discussed, feel more in control, and better understand their health. Generally, patients did not feel they would worry more, and 49% (44/90) felt they would not find the notes too confusing (although 32/90, 35% were not sure, and 14/90, 16% felt they would find the notes confusing). The patients who were interviewed were mainly interested in seeing the notes out of curiosity and because they felt they have the right to see information that is written about them. Patients also noted that it is difficult to remember everything from the visit, especially after receiving bad news, and reading the notes would help. The patients interviewed also mentioned security and privacy concerns with shared notes.

Strengths and Limitations

A major strength of this study is the use of mixed methods to gather opinions from both physicians and patients. The survey allowed us to gather opinions from a broad sample of physicians and patients, whereas the interviews allowed us to gain insight into the thoughts behind the responses. This gives us a good picture of the current mindset of these two major stakeholder groups. Another strength is the broad sample of participants, with physicians from both an academic and a nonacademic hospital and a variety of departments. However, this study had some important limitations. The survey that we used was not validated; to our knowledge, no validated survey exists on this subject. We created a survey based on the literature and pilot tested it before deploying it, ensuring that the survey questions were clear and complete according to our pilot participants. We cannot determine the response rate because we do not know how many physicians were invited or how many read the invitation email. To ensure anonymity, we did not attempt to prevent the same person from filling in the survey multiple times, although we saw no evidence of this. We also did not document any information about patients who declined to participate in the survey. The age and gender of the physicians who responded were approximately similar to the demographics of physicians in the participating hospitals. The patients who participated were more likely to be female, which might be due to a higher percentage of women in some clinics (eg, gynecology), a general participation bias (as women are generally more likely to participate in surveys or studies [18]) or may be due to a participation bias in patients who perceived themselves as similar to the researcher who distributed the surveys (who is also female). We did not ask patients about their medical conditions, and waiting rooms were shared between several outpatient clinics. Thus, our method should provide some variety in the medical conditions of patient participants, but we do not know how much. Physicians and patients with strong feelings about shared notes may be more likely to participate. The fact that all the doctors who were interviewed had a predominantly negative impression of shared notes suggests a participation bias in the interviews. Another potential source of bias is that one researcher invited patients to the surveys, entered data from paper surveys, conducted the interviews, and performed the transcribing and coding. However, a sample of the interviews and coding was checked by a second researcher to reduce the risk of bias. No field notes were made during the interviews, and the transcripts were not checked by the participants. Finally, the choice of hospitals was based on convenience, and therefore, the responses might not be representative of all Dutch hospitals. However, we included physicians from 2 centers, one academic and one nonacademic, and succeeded in including participants from a broad sample of departments.

Comparison With Previous Work

Two previous studies have investigated clinicians' opinions before the implementation of shared notes; both were focused on psychiatric care, one in the United States [10] and one in Sweden [11]. In contrast to our study, 82% of participants in the US study were positive about shared notes [10]. The Swedish study did not explicitly ask participants if they wanted to share their notes [11]. Participants in both of the aforementioned studies expressed concerns similar to those in our study: causing unnecessary worry for the patient (77% and 58%, respectively), being more confusing than helpful (67% and 53%, respectively), spending more time answering questions outside of visits (46% [10]) or being contacted with questions (69% [11]), and details being omitted from the notes (69% [10]) or being less candid in the documentation (42% [11]). In addition to the issues raised in previous studies, our physicians also expressed concern that additional time needed for documentation.

One previous study gathered patient opinions before implementation in ophthalmology patients in the United States [19]. Similar to our patients, those in this earlier study were positive about shared notes (95%). Patients felt that it would help them to better: understand their conditions (95%), remember their care plan (94%), feel more in control (90%), be prepared for visits (89%), and take better care of themselves (84%). Unlike our patients, patients in this study also believed it would help them to take their medications (77%) and rated their own anticipated ability to understand the notes as 7.5 out of 10. Studies conducted after the implementation of shared notes have found that the perceived benefits and concerns were similar to those found before implementation in both clinicians [20-23] and patients [6,7,9,12,24-27]. However, all outcome measures in these studies were assessed subjectively, with the exception of Ross and Lin [21], who found that the number of messages from the patient to the doctor increased by 31% after the implementation of shared notes. Thus, for the most part, we still do not know if the concerns raised in our study are likely to manifest or if the perceived benefits will be realized if shared notes are implemented.

Interpretation and Implications

One important finding in our study is that many patients expect the note to be written in layman's terms, whereas many physicians do not want to change the way their notes are written to make them more understandable to patients. This mismatch of expectations must be addressed if the benefits of shared notes are to be realized—patients must understand the notes in order for them to have any benefit. However, clinical jargon exists because it is a precise and efficient language for physicians to document findings and communicate with colleagues. Physicians are rightfully concerned that having to include a plain-language explanation of jargon terms with every clinical note would increase documentation time, ultimately adversely affecting patient care. However, a possible solution to this could be the automated interpretation of clinical notes. van Mens et al [28] have reported promising results in their efforts to translate diagnoses to layman's terms using SNOMED-CT (Systematized Nomenclature of Medicine-Clinical Terms); similar technology could be used to produce an explanation in layman's terms while still allowing physicians to communicate effectively with one another.

Another major concern raised by physicians is the need to document sensitive information. This is supported by Erlingsdóttir et al [29], who also reported concerns about patient privacy and confidentiality in their analysis of 1554 free-text answers from two web surveys conducted among health care providers in Sweden. Examples raised by our participants included the need to document communication with a minor in situations where the parents have mental health issues, the need to document cases of suspected abuse, and the need to document problems that the patient themselves has not yet accepted. Physicians were also concerned about how patients would respond to reading the differential diagnoses, which often contain some worrisome possibilities. Physicians feared that this important information would either be documented in other parts of the record, making it more likely to be missed on subsequent visits, or simply not be documented at all, which poses serious risks for patient care.

Another potential issue raised by our participants was the notes written by trainees. This is supported by Kung et al [30], who found that 20% of notes written by trainees raised some concerns. Trainees may be more likely to inadvertently use language that is offensive to patients. As part of the learning process, trainees must create a differential diagnosis list. The differential is the part of the notes that our physicians were most worried about sharing, as it often contains at least some alarming (although usually unlikely) possibilities. A possible solution is to document the differential and trainee notes in another part of the record; however, this runs the risk of fragmenting the record and making information more difficult to find. Another risk is that the visit note effectively becomes a note only for patients, with only a cursory summary, and that the "real" notes simply move to another field in the record.

For the most part, the findings from our patient survey and interviews were in line with previous research. In addition to the questions drawn from previous surveys, we asked our participants what they would do if reading the notes raised questions. They indicated that they would most likely search on the internet or ask at the next appointment; only a minority indicated that they would call the clinic or send a message via the patient portal. This may indicate that the increase in workload resulting from sharing visit notes would be manageable. Our patients also raised concerns about security and privacy, both in the technical sense and socially (eg, whether informal caregivers also have access to the notes).

These findings are important for hospitals seeking to implement shared notes, both in the Netherlands and elsewhere. In the Netherlands, the implementation of shared notes would consist of releasing the notes to patients in the patient portal. The concerns raised by the physicians and patients in our study should be investigated and addressed before implementation is attempted. Care should be taken to sincerely address these concerns to avoid maladaptive responses, such as moving clinical documentation to other parts of the record. Particular attention should be paid to departments who have pediatric and adolescent patients, especially in situations where giving parents access to the record may lead to harm to the patient. Differences in the content of visit notes between departments should also be considered, as well as differences between patients (eg, patients with chronic diseases may understand more of the jargon about their disease than patients with acute disease). We should also take note of the benefits that patients and physicians see in sharing the notes and find a solution that best delivers these benefits while avoiding the pitfalls foreseen by our participants.

Future research should investigate these possible solutions, preferably with the measurement of objective outcomes alongside subjective outcomes. Some important outcomes are inherently subjective, such as patients' trust in the health care system and sense of empowerment. However, the effects on communication and workflow can and should be measured objectively, such as the time needed for documentation, the ability of other physicians to find needed information, and patients' understanding of their medical situation. Future work should also repeat some of the questions presented in our survey but with an example of a visit note, so that patients are better able to say whether they are interested in the content of the note and can understand it. Future studies could also explore the relationship between factors such as health status and interest in and perceived benefits of open notes. Patients with poor health may have less energy to read notes or may be even more interested in their notes than patients with better health.

Conclusions

This mixed methods study investigated patients' and physicians' opinions of shared visit notes in the outpatient clinic setting in the Netherlands. Patients generally favored sharing notes (70/90, 77%), whereas physicians were often opposed (282/345, 81.7%). We found a mismatch between patients' and physicians' expectations for the language used in clinical notes; patients expected notes in layman's terms, whereas physicians need to communicate using precise clinical terms. Physicians raised concerns about documenting sensitive information, worrying patients with clinical suspicions and the differential diagnosis, and poorer communication due to fragmenting of the clinical documentation; patients raised concerns about security and privacy. Patients and a minority of physicians saw potential benefits in providing patients with better insight into their health state and better retention of important information from the patient visit. Hospitals seeking to implement shared notes should investigate and address these concerns, and future work should measure the effects of shared notes (objectively, when possible) to better understand if the concerns manifest as problems and if the anticipated benefits are realized.

Acknowledgments

No external funding was received for this study. The authors would like to thank MG Ros for his assistance and input during this work. They want to thank the colleagues of the EvA Servicecentrum for their support during this work, especially the eHealth team. The authors would also like to thank Dr M Janssen for distributing the physician invitations and the physicians and patients who participated in this study.

Authors' Contributions

The study design and data collection were performed by SLJ and SM. Access to patients and physicians was arranged by NVT. Data analysis was performed by SLJ. NVT and SM supervised the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Physician and patient survey instruments.
[DOCX File , 34 KB - jmir_v23i8e27764_app1.docx]

References

1. Health Promotion Glossary. World Health Organization. 1998. URL: https://apps.who.int/iris/handle/10665/64546 [accessed 2020-06-20]
2. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academies Press; 2001.
3. Elektronische Gegevensuitwisseling in De Zorg: De Wet Cliëntenrechten Bij Elektronische Verwerking Van Gegevens in De Zorg [electronic Data Interoperability in Health Care: the Clients' Rights Law in Electronic Processing of Data in Health Care]. Dutch Ministry of Health. 2017. URL: https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/brochures/2017/06/01/elektronische-gegevensuitwisseling-in-de-zorg/Wet+elektronische+verwerking+van+gegevens+20170620.pdf [accessed 2020-06-20]
4. Leveille SG, Walker J, Ralston JD, Ross SE, Elmore JG, Delbanco T. Evaluating the impact of patients' online access to doctors' visit notes: designing and executing the OpenNotes project. BMC Med Inform Decis Mak 2012 Apr 13;12:32 [FREE Full text] [doi: 10.1186/1472-6947-12-32] [Medline: 22500560]
5. Delbanco T, Walker J, Bell SK, Darer JD, Elmore JG, Farag N, et al. Inviting patients to read their doctors' notes: a quasi-experimental study and a look ahead. Ann Intern Med 2012 Oct 2;157(7):461-470 [FREE Full text] [doi: 10.7326/0003-4819-157-7-201210020-00002] [Medline: 23027317]
6. Woods SS, Schwartz E, Tuepker A, Press NA, Nazi KM, Turvey CL, et al. Patient experiences with full electronic access to health records and clinical notes through the My HealtheVet Personal Health Record Pilot: qualitative study. J Med Internet Res 2013;15(3):e65 [FREE Full text] [doi: 10.2196/jmir.2356] [Medline: 23535584]
7. Nazi KM, Turvey CL, Klein DM, Hogan TP, Woods SS. VA OpenNotes: exploring the experiences of early patient adopters with access to clinical notes. J Am Med Inform Assoc 2015 Mar;22(2):380-389. [doi: 10.1136/amiajnl-2014-003144] [Medline: 25352570]
8. Peck P, Torous J, Shanahan M, Fossa A, Greenberg W. Patient access to electronic psychiatric records: a pilot study. Health Policy Technol 2017;6:309-315 [FREE Full text] [doi: 10.1016/j.hlpt.2017.06.003]
9. Cromer R, Denneson LM, Pisciotta M, Williams H, Woods S, Dobscha SK. Trust in mental health clinicians among patients who access clinical notes online. Psychiatr Serv 2017 May 1;68(5):520-523 [FREE Full text] [doi: 10.1176/appi.ps.201600168] [Medline: 28142383]
10. Dobscha SK, Denneson LM, Jacobson LE, Williams HB, Cromer R, Woods S. VA mental health clinician experiences and attitudes toward OpenNotes. Gen Hosp Psychiatry 2016;38:89-93. [doi: 10.1016/j.genhosppsych.2015.08.001] [Medline: 26380876]
11. Petersson L, Erlingsdóttir G. Open notes in swedish psychiatric care (part 1): survey among psychiatric care professionals. JMIR Ment Health 2018 Feb 2;5(1):e11 [FREE Full text] [doi: 10.2196/mental.9140] [Medline: 29396386]
12. Moll J, Rexhepi H, Cajander, Grünloh C, Huvila I, Hägglund M, et al. Patients' experiences of accessing their electronic health records: national patient survey in Sweden. J Med Internet Res 2018 Nov 1;20(11):e278 [FREE Full text] [doi: 10.2196/jmir.9492] [Medline: 30389647]
13. Communicatie Met Patiënten via Het Patiëntenportaal [communication With Patients via the Patient Portal]. Schreinemakers B. 2018. URL: https://www.nfu-ehealth.nl/documenten/13.50-Communicatie-met-patiënten-via-het-patiëntenportaal.pdf [accessed 2020-06-20]
14. Klein JW, Jackson SL, Osteer NV, Peacock S, Delblanco T, Walker J, et al. Who reads their doctor’s notes? Examining the association between preconceptions and accessing online clinical notes. Journal of AHIMA 2018 Jan 1:28-33 [FREE Full text]
15. Fossa AJ, Bell SK, DesRoches C. OpenNotes and shared decision making: a growing practice in clinical transparency and how it can support patient-centered care. J Am Med Inform Assoc 2018 Sep 1;25(9):1153-1159. [doi: 10.1093/jamia/ocy083] [Medline: 29982659]
16. Gerard M, Chimowitz H, Fossa A, Bourgeois F, Fernandez L, Bell SK. The importance of visit notes on patient portals for engaging less educated or nonwhite patients: survey study. J Med Internet Res 2018 May 24;20(5):e191 [FREE Full text] [doi: 10.2196/jmir.9196] [Medline: 29793900]
17. Westerman MJ, Hak T, Sprangers MA, Groen HJ, van der Wal G, The A. Listen to their answers! Response behaviour in the measurement of physical and role functioning. Qual Life Res 2008 May;17(4):549-558 [FREE Full text] [doi: 10.1007/s11136-008-9333-6] [Medline: 18389384]
18. Søgaard AJ, Selmer R, Bjertness E, Thelle D. The Oslo Health Study: the impact of self-selection in a large, population-based survey. Int J Equity Health 2004 May 6;3(1):3 [FREE Full text] [doi: 10.1186/1475-9276-3-3] [Medline: 15128460]
19. Lee BS, Oster NV, Chen GY, Ding LL, Walker JD, Elmore JG. Ophthalmology patients' interest in online access to clinic notes at three US clinics. Ophthalmic Physiol Opt 2017 Dec;37(4):420-427. [doi: 10.1111/opo.12367] [Medline: 28417473]
20. Petersson L, Erlingsdóttir G. Open notes in swedish psychiatric care (part 2): survey among psychiatric care professionals. JMIR Ment Health 2018 Jun 21;5(2):e10521 [FREE Full text] [doi: 10.2196/10521] [Medline: 29929946]
21. Ross S, Lin CT. A randomized controlled trial of a patient-accessible electronic medical record. In: AMIA Annu Symp Proc. 2003 Presented at: AMIA'03; June 8-9, 2003; New York, USA p. 990 URL: http://europepmc.org/abstract/MED/14728493

22. Earnest MA, Ross SE, Wittevrongel L, Moore LA, Lin CT. Use of a patient-accessible electronic medical record in a practice for congestive heart failure: patient and physician experiences. J Am Med Inform Assoc 2004;11(5):410-417 [FREE Full text] [doi: 10.1197/jamia.M1479] [Medline: 15187074]
23. Denneson LM, Chen JI, Pisciotta M, Tuepker A, Dobscha SK. Patients' positive and negative responses to reading mental health clinical notes online. Psychiatr Serv 2018 May 1;69(5):593-596. [doi: 10.1176/appi.ps.201700353] [Medline: 29493408]
24. Wolff JL, Darer JD, Berger A, Clarke D, Green JA, Stametz RA, et al. Inviting patients and care partners to read doctors' notes: OpenNotes and shared access to electronic medical records. J Am Med Inform Assoc 2016 Aug 7:e166-e172. [doi: 10.1093/jamia/ocw108] [Medline: 27497795]
25. Kayastha N, Pollak KI, LeBlanc TW. Open oncology notes: a qualitative study of oncology patients' experiences reading their cancer care notes. J Oncol Pract 2018 Apr;14(4):e251-e258. [doi: 10.1200/JOP.2017.028605] [Medline: 29443650]
26. Mishra VK, Hoyt RE, Wolver SE, Yoshihashi A, Banas C. Qualitative and quantitative analysis of patients' perceptions of the patient portal experience with opennotes. Appl Clin Inform 2019 Jan;10(1):10-18 [FREE Full text] [doi: 10.1055/s-0038-1676588] [Medline: 30602196]
27. Denneson LM, Cromer R, Williams HB, Pisciotta M, Dobscha SK. A qualitative analysis of how online access to mental health notes is changing clinician perceptions of power and the therapeutic relationship. J Med Internet Res 2017 Jun 14;19(6):e208 [FREE Full text] [doi: 10.2196/jmir.6915] [Medline: 28615152]
28. van Mens HJ, de Keizer NF, Nienhuis R, Cornet R. Clarifying Diagnoses to Laymen by Employing the SNOMED CT Hierarchy. Stud Health Technol Inform 2018;247:900-904. [Medline: 29678091]
29. Erlingsdóttir G, Petersson L, Jonnergård K. A theoretical twist on the transparency of open notes: qualitative analysis of health care professionals' free-text answers. J Med Internet Res 2019 Sep 25;21(9):e14347 [FREE Full text] [doi: 10.2196/14347] [Medline: 31573905]
30. Kung S, Lapid MI, Swintak CC, Agarwal Z, Ryan SM, Lineberry TW. Survey of sensitive information written in patient notes by psychiatry trainees. Acad Psychiatry 2013 Jan 1;37(1):31-34. [doi: 10.1176/appi.ap.11110195] [Medline: 23338870]

Abbreviations

PHP: PHP: Hypertext Preprocessor
SNOMED-CT: Systematized Nomenclature of Medicine-Clinical Terms

Edited by R Kukafka; submitted 05.02.21; peer-reviewed by G Hailiye, L Petersson, A Roy, J Marconato; comments to author 04.04.21; revised version received 21.05.21; accepted 15.06.21; published 11.08.21.

Please cite as:
Janssen SL, Venema-Taat N, Medlock S
Anticipated Benefits and Concerns of Sharing Hospital Outpatient Visit Notes With Patients (Open Notes) in Dutch Hospitals: Mixed Methods Study
J Med Internet Res 2021;23(8):e27764
URL: https://www.jmir.org/2021/8/e27764
doi:10.2196/27764
PMID:34383660

©Sharon L Janssen, Nynke Venema-Taat, Stephanie Medlock. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 11.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Effects of Patient Portal Use on Patient Satisfaction: Survey and Partial Least Squares Analysis

Aaron P Kinney[1], DBA; Balaji Sankaranarayanan[2], PhD

[1]Herma Heart Institute, Children's Wisconsin, Milwaukee, WI, United States

[2]Department of Information Technology and Supply Chain Management, College of Business and Economics, University of Wisconsin - Whitewater, Whitewater, WI, United States

Corresponding Author:
Balaji Sankaranarayanan, PhD
Department of Information Technology and Supply Chain Management, College of Business and Economics, University of Wisconsin - Whitewater
800 W. Main Street
Whitewater, WI, 53190
United States
Phone: 1 262 472 5467
Email: sankarab@uww.edu

Abstract

Background: With digital delivery of health care services gaining prominence, patient portals have become a mainstay of many health care organizations. Despite the importance of patient portals, inconclusive data exist regarding the effect of patient portal use on patient satisfaction.

Objective: The aim of this study is to understand the relationship between the postadoptive use of patient portals and patient satisfaction outcomes.

Methods: Postadoptive use of patient portals has a positive relationship with the 3 dimensions of patient satisfaction, mediated by gratification, health self-awareness, and health perceptions. A total of 504 valid patient portal user responses were collected, and partial least squares analysis was performed to analyze the data.

Results: Patient satisfaction was captured using three dimensions: care team interaction, atmosphere, and instruction effectiveness. The results show that postadoptive use of patient portals has a positive influence on all 3 dimensions of patient satisfaction through the mediating variables of gratification, health self-awareness, and health perceptions. Specifically, postadoptive use had significant positive influence on gratification, health self-awareness, and health perceptions. Each of the 3 patient perceptions had significant positive influence on all 3 dimensions of patient satisfaction: care team interaction, atmosphere, and instruction effectiveness. Specifically, our model explained 31.8% of the care team interaction, 40.6% of the atmosphere, and 39.1% of the instruction effectiveness.

Conclusions: Our model shows that patient portal use can influence patient satisfaction through the mediating effects of gratification, health self-awareness, and health perception. Patient satisfaction is an important outcome for health care organizations. Therefore, by promoting effective patient portal use and fostering patient perceptions, health care organizations can improve patient satisfaction.

(J Med Internet Res 2021;23(8):e19820) doi:10.2196/19820

KEYWORDS

patient portal; patient satisfaction; gratification; health self-awareness; post-adoptive use; health perceptions

Introduction

Background

Increasingly, as health care digital services have gained prominence, patient portals have become a mainstay of many health care organizations. The impetus for implementing patient portals can be traced to the Affordable Care Act and Health Information Technology for Economic and Clinical Health Act, which tied reimbursements to the implementation of health information infrastructure and achievement of benchmark patient satisfaction scores [1]. Practitioner reports state that 90% of the health care organizations had implemented a patient portal in some form by 2018 [2]. The portals' services range from providing offline health records, messaging, and alerts to providing web-based real-time doctor consultations [3]. However, this leads to the key research question: does a patient's use of a portal lead to increased patient satisfaction?

As a performance metric, patient satisfaction scores are extremely important for health care organizations. Patient satisfaction scores, also known as Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) scores [4], affect the payments made to health care organizations [1]. In addition, clinical benefits such as improved clinical outcomes and improvement in overall clinical care constitute a significant benefit of higher patient satisfaction [4]. Although satisfied patients gain these significant benefits, dissatisfied patients can ignore—or worse, completely abandon—the care provided. As patient portals continue to gain prominence, improved patient satisfaction can enable organizations to provide effective health services through patient portals. Thus, studying the link between patient portal use and patient satisfaction becomes an extremely interesting and important research problem [5].

Despite strong academic interest in patient satisfaction as a performance metric, up to 45% of the current studies on the effect of portals on patient satisfaction have either shown no effect or have been inconclusive [6]. Although patient portals have been found to be associated with better patient retention [7], increased patient care compliance [8], reduction of medication errors [9], and improvement in communication [10], these relationships do not seem to have a direct link to the HCAHPS dimensions of patient satisfaction. Although prior findings on the benefits of patient portals are unequivocal, the relationship between portal use and patient satisfaction is still unclear.

Our study seeks to address this research gap in 2 ways. We theorized and tested a research model linking patient portal use and the *HCAHPS dimensions* of patient satisfaction. Furthermore, to address the research gap in our understanding of the link between portal use and satisfaction, we theorized on the mediators that link portal use and patient satisfaction. Using the Adaptive Structuration Theory (AST) and prior studies, we modeled patient perceptions as mediators of the relationship between patient portal use and patient satisfaction. In the following section, we elaborate on the prior studies on patient portals. Next, we describe how we built and tested a research model of patient portal use, patient perceptions, and patient satisfaction.

Prior Work

Our literature review revealed several important themes regarding patient portal use and its impacts. Some key articles on patient portals and their impacts are presented in Multimedia Appendix 1 [1,3,6,7,9-20]. The factors that predict the adoption and use of patient portals have received extensive attention in the literature. The Technology Acceptance Model (TAM) and the Unified Theory of Acceptance and Use of Technology (UTAUT) offer interesting theoretical perspectives to study the implementation of patient portals. Research applying this framework shows that perceived value and ease of use are important predictors of portal adoption [5-8]. The rationale is that patient portals offer the ability to access data, scheduling, and reports at any time (perceived usefulness), in addition to allowing certain functionalities without any of the previous barriers (ease of use), such as scheduling an appointment without waiting on a callback. Recent research also acknowledges that adoption is different from use and that adoption may not necessarily translate into the use of a patient portal [9].

Although the UTAUT and TAM provide an excellent framework to examine the phenomena of *adoption*, there is a significant gap in this research regarding the impacts of the system *after* use, that is, postadoptive use. Several studies have attempted to address this research gap by focusing on the relationship between patient portal use and patient outcomes. The portal benefits of improving care and communication between patients and providers [10,11], discovering medical errors, and ensuring that patients take medication on time [12] have been documented in the literature. However, findings from review studies also confirm the inconsistency in patient satisfaction studies [6]. Although some studies have found improved patient satisfaction when electronic patient access to medical information was used [6,13,14,21], other studies have found mixed results [7,15,22], and a few have found that patient satisfaction data remained unchanged [23,24].

Two important aspects of prior research on patient satisfaction must be mentioned. First, although using HCAHPS measures are considered the standard way to measure patient satisfaction at health care organizations, patient satisfaction has not been measured consistently, or it has been measured using only 1 or 2 HCAHPS measures [14,16]. This implies that patient satisfaction needs to be studied more comprehensively by enumerating its underlying dimensions. This will provide a richer description of the influence of patient portals.

Second, because of the research gap in the literature regarding the link between patient portal use and patient satisfaction, the mediators linking these factors need to be considered. Prior research suggests that patient perceptions could play a mediating role in determining patient satisfaction through patient portal use [7,14]. Patient portal use can influence a positive patient experience for patients [3]. Patients with chronic disease have mixed attitudes regarding patient portal use [17]. This suggests that studying the mediating role of patient perceptions could help address the link between patient portal use and patient satisfaction.

Goal of This Study

Prior studies show that patient satisfaction and patient portals have many important intertwined relationships to be explored. Although patient portals have been explored in terms of their impact on different aspects of the care continuum, there remains an inconsistent understanding of the relationship between portal use and patient satisfaction. As explained earlier, an understanding of the mediating influences of patient perceptions will shed light on the links between patient portals and patient satisfaction. Therefore, the 2 research questions addressed in this study are as follows:

1. What are the influences of patient portal use on patient satisfaction?

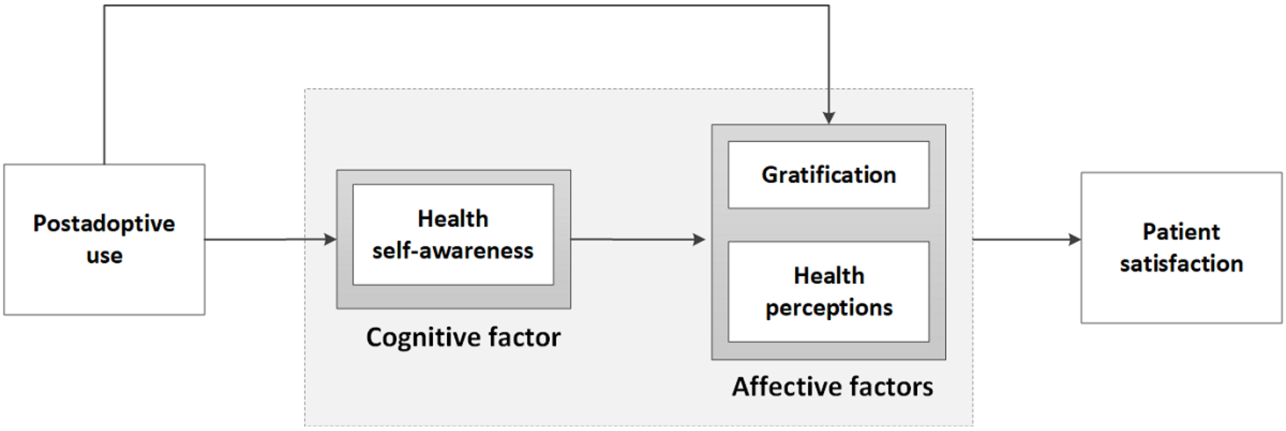
2. What are the influences of patient portal use on patients' perceptions of their own health and portal?

Research Model and Hypotheses Development

The AST [25] serves as the theoretical perspective for our research model. The AST has been applied to study adoption perceptions and behaviors in a variety of contexts such as group decision support systems and enterprise systems. Drawing upon the structuration theory formulated by Giddens [26], the AST offers a theoretical framework that delves into the dynamic relationship that connects *structures* provided by technology and the ways in which these structures are *appropriated or adapted* by a user. Traditionally, the AST had been applied in the organizational or group context [25]. However, recent research has adapted this framework to the individual level [27]. Applying the AST to the individual level, the input-process-output framework of the AST can be explained as follows: (1) input: technology in use affords certain features or structures; (2) process: technology is adapted by users to accomplish a task, and users develop certain attitudes, perceptions, and behaviors; and (3) output: processes can influence decision-making or performance outcomes [27].

Adapting the AST's input-process-output framework in our study, we posited that (1) patient portals afford certain technology features for use, (2) portal use can influence patient perceptions, and (3) patient perceptions can influence patient satisfaction outcomes. The research model of patient portal use and patient satisfaction is presented in Figure 1. The model shows that postadoptive use will have a positive influence on health self-awareness, gratification, and health perceptions. We classified them as cognitive factor (knowledge-based) and affective factor (emotion-based). These patient perceptions are posited to have a positive influence on patient satisfaction.

Figure 1. Influences of postadoptive use of electronic portals on patient perceptions and satisfaction.



Dimensions of Patient Satisfaction

As noted earlier, patient satisfaction has not been conceptualized or measured consistently as an outcome variable in prior research. It is common for numerous measures of patient satisfaction to be used, including likeliness to recommend [28], satisfaction with nursing [29], and satisfaction with physician communication [30]. Some studies have only used 1 or 2 items from the HCAHPS survey (either overall satisfaction or willingness to recommend) to measure patient satisfaction [14,16].

We diverged from these studies to conceptualize patient satisfaction as a multidimensional construct using multiple measures. The conceptualization and measures used in this study for patient satisfaction were based on the HCAHPS survey. This is the national survey used by health care organizations to capture patient perceptions of hospital experience and is used by the Centers for Medicare & Medicare Services to standardize medical reimbursement [4]. This survey is typically used to gauge patient satisfaction at hospitals [31] and has been a critical component in bringing about transparency to patient perceptions of care [32]. On the basis of the HCAHPS survey, patient satisfaction was theorized in this study as consisting of three dimensions: care team interaction, atmosphere, and instruction effectiveness. Care team interaction refers to any communication between a patient and a member of the care team, such as providers and nurses. Atmosphere refers to the evaluation of the health encounter with respect to the environment around the patient, notably cleanliness, quietness, and staff responsiveness. Instruction effectiveness captures the care team's ability to convey pertinent information to the patient, including communication about medicines and discharge information.

Health Self-awareness

In this study, health self-awareness was defined as the extent of knowledge and skill sets of patients in relation to their own health. Prior research has shown that patients are more informed about their health through patient portal use [11,33]. As patients continue to use patient portals, these technology features [25] provide them with access to accurate information about their health status [34]. In addition, patient portals provide detailed explanations of the test results and associated health conditions [19]. Patients can therefore compare such information (eg, blood work results) against established benchmarks, in addition to obtaining a detailed understanding of their health condition. As patients accumulate knowledge about their health, they are more likely to be informed and involved in their health care decisions [35]. Informed patients can proactively request specific health services from their providers [36], which can reduce medical errors [12,34], improve decision-making [37], and subsequently influence the 3 dimensions of patient satisfaction (care team interaction, atmosphere, and instruction effectiveness). Hence, we postulated the following hypotheses:

- H_1: Health self-awareness mediates the positive influence of postadoptive use on patient satisfaction.
- H_{1a}: Postadoptive use will have a positive influence on health self-awareness.
- H_{1b}: Health self-awareness will have a positive influence on patient satisfaction.

Gratification

Gratification has been defined in numerous ways, particularly in relation to the Uses and Gratification Theory [38,39]. In this study, we defined gratification as a feeling of pleasure directly related to achieving a desired task such as scheduling, reviewing medical information, or using the patient portal. Patient portal use provides a near-instant ability to achieve health-related tasks [40,41], rather than limiting individuals to accomplishing such tasks only during business hours [42]. As the goal of portal use is information seeking or knowledge gathering [41], the immediate sharing of information is likely to gratify the patient. Gratified patients will reflect positively upon their care experience [41,43], which will have a positive influence on their satisfaction with overall health care delivery. Hence, we posited the following:

- H_2: Gratification mediates the positive influence of postadoptive use on patient satisfaction.
- H_{2a}: Postadoptive use will have a positive influence on gratification.
- H_{2b}: Gratification will have a positive influence on patient satisfaction.

Health Perceptions

In this study, health perceptions were defined as having positive emotions about a person’s health. Prior work in psychology literature has shown the importance of positive emotions as they improve creative problem solving [44] and satisfaction [45] and increase the likelihood of success [46]. Research in information systems literature has also pointed to the centrality of positive emotions in predicting the use of systems. For instance, knowledge gained through the UTAUT [47] pointed to affect and associated constructs such as computer playfulness as antecedents of behavioral intention to use a system.

We posited that health perceptions would mediate the relationship between postadoptive use and patient satisfaction. First, prior studies have informed us that health perceptions can influence outcome variables, including practice satisfaction in physicians [45,46,48]. As users gather information, they experience positive health perceptions, and this affective attitude component directly contributes to their satisfaction attitudes. As long as the information technology system (ie, portal) continues to enable this behavior [25], it has the potential to allow users to feel more positive about their health. As patients have more positive feelings about their health status, they are more likely to rate the status of their health experience higher, leading to higher patient satisfaction scores. Hence, we posited the following:

- H_3: Health perceptions mediate the positive influence of postadoptive use on patient satisfaction.
- H_{3a}: Postadoptive use will have a positive influence on health perceptions.
- H_{3b}: Health perceptions will have a positive influence on patient satisfaction.

Indirect Effects of Health Self-awareness

Health self-awareness can have a positive impact on gratification and health perceptions, thereby indirectly influencing patient satisfaction. On the basis of social cognitive theory [49], prior research in different contexts has shown that cognitive factors can influence affective factors. For example, in a learning context, working memory skills (cognitive factor) can influence writing anxiety and self-efficacy (affective factors), further affecting writing performance [50]. Similarly, in a health care context, mindfulness, a cognitive factor, can have a positive impact on affective empathy, leading to improved engagement in nursing [51]. As explained earlier, as a patient’s knowledge increases, they are more likely to be informed and involved in their own health care decisions [35]. As patients begin to make informed decisions, it can lead to a higher level of gratification (ie, satisfaction with learning through technology use) and health perceptions (ie, a positive feeling about taking control of their own health). On the other hand, as health self-awareness decreases, the patients’ capacity to make informed decisions also decreases, which can lead to lower levels of gratification and health perceptions. Hence, we posited the following:

- H_4: Health self-awareness will have an indirect influence on patient satisfaction through gratification and health perceptions.
- H_{4a}: Health self-awareness will have a positive influence on gratification.
- H_{4b}: Health self-awareness will have a positive influence on health perceptions.

Methods

Recruitment

As part of a larger research study, this survey was designed as a nationwide electronic survey to be disseminated within the United States. This study used a Qualtrics panel (Qualtrics) with a financial incentive provided to the respondents (the negotiated rate with Qualtrics was a little less than US $5 per respondent). Qualtrics is a highly reputed experience management company that provides a platform for survey design and execution. It maintains a panel of respondents and recruits them depending on the purpose of the survey. For our study, we sought respondents who had visited their regular health care facility and used an electronic patient portal within the last 12 months. Respondents were excluded from participating in the survey if they did not meet these 2 criteria. Regular health care facilities were defined in the survey as health care facilities (eg, primary care provider’s office and hospital) that the respondents typically visit for health care services. The survey was expected to take approximately 15 minutes to complete.

Electronic patient portals were defined as the secure websites of the regular health care facilities that provide patients with convenient 24-hour access to their personal health information such as recent or upcoming medical visits, prescriptions, and vaccinations. Furthermore, the instructions also stated that participation in this research was anonymous and voluntary. The exclusion criteria also included respondents failing to consent to take the survey, failing to confirm that they were aged above 18 years, or failing to complete the survey. A final sample of 504 responses was obtained.

Item Development and Expert Review

Item development for this study began by using established measures. When the established measures could not be used or were modified to the extent that the expert review suggested that they be treated as new items, new items were established in a rigorous process (Textbox 1). Prior research has laid clear blueprints for item development for this survey [52,53]. For each new item, initial development was informed by a literature review and the close collaboration with an experienced information technology professor with experience in health care research. A panel of subject matter experts related to the topic at hand was then created, including physicians, nurses, health care administrators, and health care researchers. After gaining insight from the expert review and making adjustments, a pilot study was completed, consisting of 20 health care staff members, 9 of whom completed the survey. These results were used to modify and finalize the survey instruments. Subsequently, an additional pilot survey was conducted using 43 surveys completed by doctoral students.

Textbox 1. Measures and sources.

Patient satisfaction (17 Items)

- Adapted from [4]

Postadoptive portal use (4 Items)

- Adapted from [54]

Health self-awareness (3 Items)

- Adapted from [54]

Gratification (3 Items)

- Developed new items based on [38]

Health perceptions (4 Items)

- Adapted from [54]

Common Method Bias

As this study measured predictor and criterion variables using the same system, time, and source [55,56], it was necessary to test for common method bias as it increases the possibility of inflated results [55,56]. Statistical tests for common method bias [56] suggested that it may not be a significant concern in this study.

Control Variables

The control variables measured in this survey included age, gender, income, education, and race. We also modeled health anxiety as a control variable. Health anxiety was captured by the item *I am very anxious about my health* on a strongly disagree to strongly agree scale. Race was modeled as a 0-1 variable, with 1 representing Caucasian. Table 1 shows the demographic characteristics of the respondents.

Table 1. Demographic characteristics of the respondents (N=504)[a].

Characteristics	Respondents, n (%)
Age (years)	
18-25	61 (12.1)
26-30	68 (13.5)
31-35	60 (11.9)
36-40	61 (12.1)
41-45	50 (9.9)
46-50	37 (7.3)
51-55	36 (7.1)
56-60	32 (6.3)
61-65	31 (6.2)
>65	66 (13.1)
Gender	
Male	159 (31.5)
Female	343 (68.1)
Education	
Eighth grade or less	15 (2.9)
Some high school but did not graduate	115 (22.8)
High school graduate or General Educational Development certificate	182 (36.1)
Some college or 2-year degree	122 (24.2)
4-year college graduate	70 (13.9)
Income (US $)	
≤25,000	107 (21.2)
25,001-50,000	150 (29.8)
50,001-75,000	116 (23)
75,001-10,000	57 (11.3)
100,001-125,000	33 (6.5)
125,001-150,000	14 (2.8)
150,001-200,000	12 (2.4)
200,001-250,000	2 (0.4)
>250,000	6 (1.2)
Race	
Caucasian	390 (77.4)
African American	53 (10.5)
Asian	12 (2.4)
American Indian or Alaska Native	8 (1.6)
Native Hawaiian or Pacific Islander	0 (0)
Mixed race	7 (1.4)
Other	26 (5.2)

[a] For each of the demographic variables, missing data constituted the remaining percentage.

Results

Overview

We used SmartPLS (version 3.3.2, SmartPLS GmbH) structural equation modeling to analyze the data. The measurement model was first examined to evaluate reliability, convergent validity, and discriminant validity. The structural model was then evaluated to test the specific hypotheses. The measurement model and structural model analysis are presented below.

Measurement Model

Confirmatory factor analysis was performed to establish the reliability and validity of the measures [57]. The measures and loadings are presented in Table 2. Composite reliabilities were above the threshold of 0.7, and all item loadings were statistically significant and above the acceptable threshold of 0.70 [58]. Table 3 shows the validation of the measurement model for the constructs in this study.

The diagonal elements (Table 3) show the square root of the average variance extracted (AVE). The second-order constructs were modeled in partial least squares following the procedure described in the study by Pavlou and El-Sawy [59]. The dimensions of care team interaction, instruction effectiveness, and atmosphere (Table 2) were modeled as first-order constructs. The model then calculates the path weights from the first-order constructs to the second-order constructs, and latent factor scores are calculated for each second-order construct. As each second-order construct is represented by a latent factor score and not by multiple items, the AVE for all second-order constructs is 1. The off-diagonal elements show correlations among the constructs. The AVE values were found to be greater than 0.5 [60]. This establishes the convergent validity of the constructs. The item-to-construct correlations for each construct were found to be less than the square root of the corresponding AVE, thus establishing discriminant validity [60].

Table 2. Constructs, measures, and loadings.

Item	Indicator	Loading	Composite reliability
GR[a,b]			0.951
GR_1	I feel satisfied when I receive information regarding my health via the electronic patient portal	0.920	
GR_2	Accomplishing health care related tasks via the electronic patient portal when it is convenient for me is satisfying (eg, scheduling appointments, checking test results, etc)	0.930	
GR_3	I feel satisfied about my experiences completing health care related tasks via the electronic patient portal (eg, scheduling appointments, checking test results, etc)	0.941	
HSA[b,c]			0.947
HSA_1	I have a good understanding of my health status	0.947	
HSA_2	I am informed regarding ideal targets for indicators of my health (eg, weight, cholesterol, blood sugar, etc)	0.917	
HSA_3	I am knowledgeable about my health status	0.942	
PAU[b,d]			0.915
PAU_1	Please select your usage frequency - scheduling appointments	0.865	
PAU_2	Please select your usage frequency - emailing my provider	0.884	
PAU_3	Please select your usage frequency - checking test results	0.845	
PAU_4	The electronic patient portal is used frequently by me	0.818	
HP[b,e]			0.924
HP_1	I lead an active and healthy life	0.846	
HP_2	I feel optimistic about my health	0.876	
HP_3	I feel satisfied with my latest health check-up results (eg, blood pressure, cholesterol, glucose levels)	0.854	
HP_4	In general, I am enthusiastic about my health	0.891	
Care team interaction[f]			
DC[g]			0.944
DC_1	How often did providers (eg, doctors) treat you with courtesy and respect?	0.918	
DC_2	How often did providers (eg, doctors) listen carefully to you?	0.919	
DC_3	How often did providers (eg, doctors) explain things in a way you could understand?	0.927	
NC[h]			0.936
NC_1	How often did nurses treat you with courtesy and respect?	0.907	
NC_2	How often did nurses listen carefully to you?	0.931	
NC_3	How often did nurses explain things in a way you could understand?	0.897	
Atmosphere[f]			
CQ[i]			0.916
CQ_1	How often were public restrooms found clean?	0.919	
CQ_2	How often was the noise level quiet during appointments?	0.919	
SR[j]			0.866
SR_1	How often did you get help from any staff as soon as you wanted it?	0.818	
SR_2	How often did you get an appointment as soon as you needed?	0.858	
SR_3	When contacting my typical health care facility with a question, I typically received an answer the same day	0.802	

Item	Indicator	Loading	Composite reliability
Instruction effectiveness[f]			
CM[k]			0.844
CM_1	How often did hospital staff describe possible side effects?	0.810	
CM_2	How often did hospital staff tell you what the medicine was for?	0.864	
CM_3	I clearly understood the purpose for taking each of my medications	0.730	
DI[l]			0.910
DI_1	Staff took my preferences into account in deciding what my health care needs would be	0.904	
DI_2	Whenever I left my typical health care facility, I had a good understanding of the things I was responsible for in managing my health	0.888	
DI_3	I received information in writing summarizing the visits and describing any symptoms or health problems to look out for	0.840	

[a] GR: gratification.

[b] Scale used for use frequency questions: 1=never to 7=multiple times a day; scale used for all other questions: 1=strongly disagree to 7=strongly agree.

[c] HSA: health self-awareness.

[d] PAU: postadoptive use.

[e] HP: health perceptions.

[f] Patient satisfaction dimensions: second-order constructs: scale used for “how often” questions: 1=never to 7=always; scale used for other questions: 1=strongly disagree to 7=strongly agree.

[g] DC: doctor communication.

[h] NC: nurse communication.

[i] CQ: cleanliness and quietness.

[j] SR: staff responsiveness.

[k] CM: communication about medicines.

[l] DI: discharge information.

Table 3. Measurement model validation[a].

Constructs	Value, mean (SD)	ATMOS[b]	CTI[c]	IE[d]	GR[e]	HSA[f]	HP[g]	PAU[h]
ATMOS	5.82 (0.98)	*1*[i]	—[j]	—	—	—	—	—
CTI	6.07 (0.99)	0.776	*1*	—	—	—	—	—
IE	5.78 (1.05)	0.787	0.696	*1*	—	—	—	—
GR	5.32 (1.41)	0.453	0.384	0.475	*0.930*	—	—	—
HSA	5.89 (1.16)	0.540	0.496	0.528	0.515	*0.925*	—	—
HP	5.19 (1.31)	0.398	0.321	0.444	0.386	0.520	*0.867*	—
PAU	3.54 (1.47)	0.076	−0.012	0.148	0.458	0.155	0.237	*0.853*

[a] Latent scores of second-order constructs (atmosphere [ATMOS], care team interaction [CTI], and instruction effectiveness [IE]) are standardized scores. Hence mean and SD are 0 and 1, respectively. However, in this table, we have provided the mean and SD of all the corresponding items for ATMOS, CTI, and IE.

[b] ATMOS: atmosphere.

[c] CTI: care team interaction.

[d] IE: instruction effectiveness.

[e] GR: gratification.

[f] HSA: health self-awareness.

[g] HP: health perceptions.

[h] PAU: postadoptive use.

[i] The diagonals show the square root of the average variance extracted. Diagonal values for second-order constructs (atmosphere, care team interaction, and instruction effectiveness) are 1 because these are modeled using latent factor scores. Off-diagonal elements show correlation among the constructs.

[j] Not applicable.

Structural Model

Figure 2 shows the structural model with all the β values and significance of the paths. Care team interaction, instruction effectiveness, and atmosphere were modeled and validated as second-order factors [59]. The items associated with the second-order construct were first checked for convergent and discriminant validity. Next, the path coefficients from the first-order constructs to the second-order constructs were checked for significance. The corresponding path coefficients were found to be significant at *P*<.001: doctor communication to care team interaction (β=.537); nurse communication to care team interaction (β=.523); cleanliness and quietness to atmosphere (β=.488); staff responsiveness to atmosphere (β=.633); discharge instructions to instruction effectiveness (β=.597); and communication about medicines to instruction effectiveness (β=.482). This established the second-order constructs, and the latest factor scores were obtained for each dimension of patient satisfaction. The model was then tested with all constructs and latent scores of the second-order constructs.

Figure 2. Structural model of influences of postadoptive use.

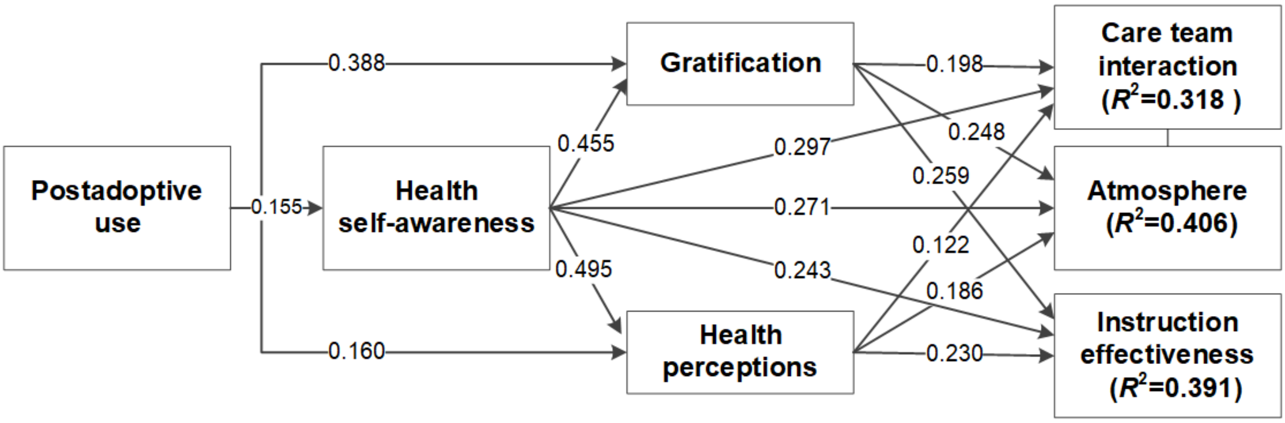


Table 4 shows the β values and *P* values for each hypothesis. All hypotheses in the research model were supported. Furthermore, the Sobel test for each mediating relationship showed that all mediating influences were statistically significant (Multimedia Appendix 2 [61]). Among the control variables, numerous significant relationships were observed. Age had a significant relationship with atmosphere (β=.209; *P*<.001), care team interaction (β=.201; *P*<.001), and instruction effectiveness (β=.110; *P*=.006). Gender was significantly related to atmosphere (β=.091; *P*=.01), care team interaction (β=.088; *P*=.03), and instruction effectiveness (β=.088; *P*=.02). Race was significantly related to atmosphere (β=.088; *P*=.02) and instruction effectiveness (β=.091; *P*=.004). Education, income, and health anxiety were not significantly related to any of the dimensions.

Table 4. Results of hypothesis testing.

Hypothesis	β	*P* value
H_{1a}		
Postadoptive use to health self-awareness	.155	<.001
H_{1b}		
Health self-awareness to care team interaction	.297	<.001
Health self-awareness to atmosphere	.271	<.001
Health self-awareness to instruction effectiveness	.243	<.001
H_{2a}		
Postadoptive use to gratification	.388	<.001
H_{2b}		
Gratification to care team interaction	.198	<.001
Gratification to atmosphere	.248	<.001
Gratification to instruction effectiveness	.259	<.001
H_{3a}		
Postadoptive use to health perceptions	.160	<.001
H_{3b}		
Health perceptions to care team interaction	.122	.009
Health perceptions to atmosphere	.186	<.001
Health perceptions to instruction effectiveness	.230	<.001
H_{4a}		
Health self-awareness to gratification	.455	<.001
H_{4b}		
Health self-awareness to health perceptions	.495	<.001

Discussion

Principal Findings

Our findings show that patient portal use has a positive influence on the three mediators: health self-awareness, gratification, and health perceptions. Each of the 3 mediators also has positive influences on the 3 dimensions of patient satisfaction: care team interaction, atmosphere, and instruction effectiveness. This study contributes to our understanding of the influence of patient portal use on patient satisfaction in 3 distinct ways. First, this study addressed an important research question regarding the link between patient portal use and patient satisfaction [25]. This study diverged from earlier studies that have used *few select items* to measure patient satisfaction by using *multiple dimensions* of the HCAHPS measures for patient satisfaction (Multimedia Appendix 3 [14,16,28,29]). This is a novel contribution to extant research because it helps us to discern the portal use influences on the various facets of patient satisfaction. Future studies can use this conceptualization of patient satisfaction not just for patient portal use, but also for other applications.

Second, this study extends prior work on the use of the TAM and UTAUT on patient portals and postadoptive use by highlighting the *pathways* through which portal use can influence patient satisfaction. We have enumerated the role played by three mediators: health self-awareness, gratification, and health perceptions. Through these mediators, this study makes novel contributions to the extant literature and practice in terms of *how* postadoptive use influences patient satisfaction. Notably, our study confirmed the role of health self-awareness as a critical mediator. This study used the AST as the underlying theory because it deals with technology adoption and use behaviors. However, psychological theories such as the Mere Exposure Effect [62-64] could also serve as a broader theoretical underpinning for our study. The Mere Exposure Effect as espoused in the study by Zajanc [62-64] states that a person's familiarity with, for example, an object would make them develop a preference for it. Applied to the context of the patient portal, a patient's familiarity with the patient portal can influence them to develop affective reactions, which can lead to patient satisfaction. In other words, efforts to design technologies that facilitate learning and knowledge acquisition can play a critical role in higher patient satisfaction scores.

Furthermore, we defined gratification as a feeling of pleasure directly related to achieving a desired task through technology use, as informed by the Uses and Gratification Theory [38]. In an era when research has demonstrated that consumers seek computer-mediated interactions and that such use can provide gratification [40], it is not surprising to find gratification playing such a critical mediating role. Our study found that as the feeling of gratification is achieved, this affective component directly affects the patient's satisfaction with the entire care experience, thereby expanding the Uses and Gratification Theory into the patient portal domain. As the patient portal is used extensively, ways to influence this variable seem critically important with regard to influencing patient satisfaction.

Finally, health perceptions were found to play a key mediating role between postadoptive use and patient satisfaction. Prior studies show that health perceptions have an impact on the satisfaction variables [45,46,48]. By empirically validating the influence of patient portal use on health perceptions, this study shows that portals have the potential to allow users to feel more positive about their health. This is an interesting finding because it suggests that information on the patient portal can emotionally engage patients. Researchers and practitioners can strive to understand the nuances in health perceptions because they can lead to higher patient satisfaction scores and potentially higher reimbursement for health care providers.

Limitations

Despite the many significant relationships discovered in this study, it includes several limitations. We used the AST as a theoretical perspective in this study. Other theoretical perspectives such as the Mere Exposure Effect could be used in future studies to build the research model. Our study included 504 survey responses collected through a cross-sectional survey. Measuring responses through a cross-sectional survey presents concerns of common method bias because the independent and dependent variables are gathered from the same respondent at the same survey session. Although common method bias was not found to be a concern in this study, it is possible that the respondents may not have fully understood the system's capabilities or may have overestimated or underestimated their use habits. In addition, survey respondents who participate in web-based surveys may be more technologically savvy than the general population. Some of the measures were adapted to the patient portal context, and some new items were created to measure the constructs. Although we followed a rigorous process of survey development and testing, specific follow-up to expand and generalize these definitions is needed. In this study, we adapted the HCAHPS measures used in US hospitals as the measure for patient satisfaction. As the scale used ranges from *never* to *always*, the relationship between the hypothesized constructs and patient satisfaction may be reflective of frequency-based measurement. Future studies can choose to pursue a different scale to measure patient satisfaction. Furthermore, the HCAHPS measures were adapted to suit the context of this study because we sought respondents who had visited their regular health care facility and used an electronic patient portal in the last 12 months. We acknowledge that patient satisfaction can be measured using different methods. Future studies can modify the existing measures or introduce new ways to measure patient satisfaction. We used health anxiety as a proxy for measuring the health status of the patient. Future studies can directly measure the emotional and physical health of the patient to use as control variables, taking into account the response bias with such measurements. Furthermore, this study did not include chronic disease status, health care use, digital literacy, and overall internet use as part of the analysis. Finally, a novel data collection approach that allows actual feedback of users' habits would greatly illuminate this discussion on patient portals.

Implications for Research

This study presents several important opportunities for future research. First, the study of gratification has an extensive history [42,65]. This research points to a pivotal role played by gratification in mediating all 3 patient satisfaction outcome variables used in this research (care team interaction, atmosphere, and instruction effectiveness). Future work combining the Uses and Gratification Theory and other constructs such as digital self-efficacy may yield important insights and help to continue the expansion of knowledge in this area.

Second, this research highlights the opportunity to take the validated survey results of self-reported data and determine a way to attempt a similar study using actual reports of user actions from the information systems themselves, rather than reports from patients. This could provide either strong confirmation of this study or yield important new research streams into how the cognitive and affective variables actually affect patient satisfaction and postadoptive use.

Third, efforts to specifically facilitate or enable health self-awareness and health perceptions offer the potential to greatly expand the understanding of patient satisfaction research. Other constructs that might have an interaction effect on postadoptive use and health self-awareness, in particular, offer the potential to shed light on key variables within the nomological network of patient satisfaction.

Implications for Practice

Our study has several important practical implications regarding the postadoptive use of patient portals, patient perceptions, and patient satisfaction. First, our study findings clearly establish a relationship between patient portal use and patient perceptions. This implies that efforts to increase patient portal enrollment by health care systems are worthwhile, especially because portal use improves the perceptions of satisfaction with the care received. There are several ways in which hospitals are rated on care provided: HCAHPS and Net Promoter scores. Therefore, any insights into improving patient perceptions of the care received become critical.

Second, from a logistics standpoint, this research suggests that patient portal use could reduce medical inefficiencies and wasted time for patients and providers alike. Patients can use the time saved by not having to make phone calls to follow up on tests, schedule appointments, and ask questions. Care providers can handle such requests more efficiently through these portals, thereby reducing errors. With the added benefit of improving patient perceptions of care provided, the mutual benefit would be remarkable.

A third practitioner implication lies in the way in which the outcome variable of patient satisfaction was measured. By measuring three factors of patient satisfaction derived directly from standard HCAHPS scoring, this research was able to identify important antecedents to these outcomes, including key variables such as instruction effectiveness. Therefore, health care organizations can focus on emerging techniques to improve their instruction effectiveness, such as coordinating with web-based apps and providing short postcheckup surveys and audio or video instructions.

Finally, given the current reality of the COVID-19 pandemic, patient portals will continue to remain a top priority for health care organizations. Our study showcases the important role played by patient portals, which enable increased exchange of health-related information without requiring face-to-face support. At this juncture, by improving the portals' capabilities and by engaging patients to use patient portals, health care organizations can enhance health choices, improving patient satisfaction in the process.

Conclusions

In this study, we sought to address a key research gap in extant studies by studying the link between patient portal use and the different dimensions of patient satisfaction. We also sought to understand the influence of patient perceptions as mediators of the link between patient portal use and patient satisfaction. Our model shows that patient portal use can influence patient satisfaction through the mediating effects of gratification, health self-awareness, and health perceptions. Future research can seek to take a more nuanced perspective on the mediators highlighted in this study. Patient satisfaction is an important outcome for health care organizations. Therefore, the findings of this study can be used by health care organizations and practitioners to promote effective patient portal use and foster patient perceptions to improve patient satisfaction.

Acknowledgments

This study benefited from the support of the Herma Heart Institute-Children's Wisconsin, Milwaukee, Wisconsin.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Key articles on patient portal use.
[PDF File (Adobe PDF File), 61 KB - jmir_v23i8e19820_app1.pdf]

Multimedia Appendix 2
Mediation analysis.
[PDF File (Adobe PDF File), 37 KB - jmir_v23i8e19820_app2.pdf]

Multimedia Appendix 3
Reflective measures model.
[PDF File (Adobe PDF File), 99 KB - jmir_v23i8e19820_app3.pdf]

References

1. Aljabri D, Dumitrascu A, Burton MC, White L, Khan M, Xirasagar S, et al. Patient portal adoption and use by hospitalized cancer patients: a retrospective study of its impact on adverse events, utilization, and patient satisfaction. BMC Med Inform Decis Mak 2018 Jul 27;18(1):70 [FREE Full text] [doi: 10.1186/s12911-018-0644-4] [Medline: 30053809]
2. Heath S. Patient portal adoption tops 90%, but strong patient use is needed. Patient Engagement HIT. 2018. URL: https://patientengagementhit.com/news/patient-portal-adoption-tops-90-but-strong-patient-use-is-needed2018 [accessed 2020-09-05]
3. Graham TA, Ali S, Avdagovska M, Ballermann M. Effects of a web-based patient portal on patient satisfaction and missed appointment rates: survey study. J Med Internet Res 2020 May 19;22(5):e17955 [FREE Full text] [doi: 10.2196/17955] [Medline: 32427109]
4. HCAHPS: Patients' perspectives of care survey. Centers for Medicare & Medicaid Services. URL: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalHCAHPS [accessed 2021-08-03]
5. Portz JD, Bayliss EA, Bull S, Boxer RS, Bekelman DB, Gleason K, et al. Using the technology acceptance model to explore user experience, intent to use, and use behavior of a patient portal among older adults with multiple chronic conditions: descriptive qualitative study. J Med Internet Res 2019 Apr 08;21(4):e11604 [FREE Full text] [doi: 10.2196/11604] [Medline: 30958272]
6. Lin C, Wittevrongel L, Moore L, Beaty BL, Ross SE. An internet-based patient-provider communication system: randomized controlled trial. J Med Internet Res 2005 Aug 05;7(4):e47 [FREE Full text] [doi: 10.2196/jmir.7.4.e47] [Medline: 16236699]
7. Hassol A, Walker JM, Kidder D, Rokita K, Young D, Pierdon S, et al. Patient experiences and attitudes about access to a patient electronic health care record and linked web messaging. J Am Med Inform Assoc 2004 Nov 01;11(6):505-513. [doi: 10.1197/jamia.m1593]

8. Tavares J, Oliveira T. Electronic health record patient portal adoption by health care consumers: an acceptance model and survey. J Med Internet Res 2016 Mar 02;18(3):e49 [FREE Full text] [doi: 10.2196/jmir.5069] [Medline: 26935646]
9. Zhong X, Park J, Liang M, Shi F, Budd PR, Sprague JL, et al. Characteristics of patients using different patient portal functions and the impact on primary care service utilization and appointment adherence: retrospective observational study. J Med Internet Res 2020 Feb 25;22(2):e14410 [FREE Full text] [doi: 10.2196/14410] [Medline: 32130124]
10. McAlearney AS, Sieck CJ, Gaughan A, Fareed N, Volney J, Huerta TR. Patients' perceptions of portal use across care settings: qualitative study. J Med Internet Res 2019 Jun 06;21(6):e13126 [FREE Full text] [doi: 10.2196/13126] [Medline: 31172960]
11. Fix GM, Hogan TP, Amante DJ, McInnes DK, Nazi KM, Simon SR. Encouraging patient portal use in the patient-centered medical home: three stakeholder perspectives. J Med Internet Res 2016 Nov 22;18(11):e308 [FREE Full text] [doi: 10.2196/jmir.6488] [Medline: 27876686]
12. 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]
13. Winstanley EL, Burtchin M, Zhang Y, Campbell P, Pahl J, Beck S, et al. Inpatient experiences with MyChart bedside. Telemed J E Health 2017 Aug;23(8):691-693 [FREE Full text] [doi: 10.1089/tmj.2016.0132] [Medline: 28328391]
14. Silver RA, Subramaniam C, Stylianou A. The impact of portal satisfaction on portal use and health-seeking behavior: structural equation analysis. J Med Internet Res 2020 Mar 27;22(3):e16260 [FREE Full text] [doi: 10.2196/16260] [Medline: 32217505]
15. Goldzweig CL, Orshansky G, Paige NM, Towfigh AA, Haggstrom DA, Miake-Lye I, et al. Electronic patient portals: evidence on health outcomes, satisfaction, efficiency, and attitudes: a systematic review. Ann Intern Med 2013 Nov 19;159(10):677-687. [doi: 10.7326/0003-4819-159-10-201311190-00006] [Medline: 24247673]
16. Hoogenbosch B, Postma J, de Man-van Ginkel JM, Tiemessen NA, van Delden JJ, van Os-Medendorp H. Use and the users of a patient portal: cross-sectional study. J Med Internet Res 2018 Sep 17;20(9):e262 [FREE Full text] [doi: 10.2196/jmir.9418] [Medline: 30224334]
17. Kruse CS, Argueta DA, Lopez L, Nair A. Patient and provider attitudes toward the use of patient portals for the management of chronic disease: a systematic review. J Med Internet Res 2015;17(2):e40 [FREE Full text] [doi: 10.2196/jmir.3703] [Medline: 25707035]
18. Dumitrascu AG, Burton MC, Dawson NL, Thomas CS, Nordan LM, Greig HE, et al. Patient portal use and hospital outcomes. J Am Med Inform Assoc 2018 Apr 01;25(4):447-453 [FREE Full text] [doi: 10.1093/jamia/ocx149] [Medline: 29300961]
19. Al-Ramahi M, Wahbeh A, Noteboome C. Features related to patient portal user satisfaction: N-gram-based analysis of users' feedback. In: Proceedings of the 2018 ACM SIGMIS Conference on Computers and People Research. 2018 Presented at: SIGMIS-CPR '18: 2018 Computers and People Research Conference; June 18 - 20, 2018; Buffalo-Niagara Falls NY USA p. 1-152. [doi: 10.1145/3209626.3209729]
20. Rozenblum R, Donzé J, Hockey PM, Guzdar E, Labuzetta MA, Zimlichman E, et al. The impact of medical informatics on patient satisfaction: a USA-based literature review. Int J Med Inform 2013 Mar;82(3):141-158. [doi: 10.1016/j.ijmedinf.2012.12.008] [Medline: 23332922]
21. Cathala N, Brillat F, Mombet A, Lobel E, Prapotnich D, Alexandre L, et al. Patient followup after radical prostatectomy by internet medical file. J Urol 2003 Dec;170(6 Pt 1):2284-2287. [doi: 10.1097/01.ju.0000095876.39932.4a] [Medline: 14634397]
22. Noteboom C, Abdel-Rahman M. Toward actionable knowledge: a systematic analysis of mobile patient portal use. In: Wickramasinghe N, Bodendorf F, editors. Delivering Superior Health and Wellness Management with IoT and Analytics. Cham: Springer; 2019:587-602.
23. Griffin A, Skinner A, Thornhill J, Weinberger M. Patient Portals: Who uses them? What features do they use? And do they reduce hospital readmissions? Appl Clin Inform 2017 Dec 16;07(02):489-501. [doi: 10.4338/aci-2016-01-ra-0003]
24. Ross SE, Moore LA, Earnest MA, Wittevrongel L, Lin C. Providing a web-based online medical record with electronic communication capabilities to patients with congestive heart failure: randomized trial. J Med Internet Res 2004 May 14;6(2):e12 [FREE Full text] [doi: 10.2196/jmir.6.2.e12] [Medline: 15249261]
25. DeSanctis G, Poole MS. Capturing the complexity in advanced technology use: adaptive structuration theory. Organ Sci 1994 May;5(2):121-147. [doi: 10.1287/orsc.5.2.121]
26. Giddens A. The Constitution of Society: Outline of the Theory of Structuration. Berkeley, California: University of California Press; 1986:1-440.
27. Schmitz KW, Teng JT, Webb KJ. Capturing the complexity of malleable it use: adaptive structuration theory for individuals. MIS Q 2016 Mar 3;40(3):663-686. [doi: 10.25300/misq/2016/40.3.07]
28. Tsai T, Orav E, Jha A. Patient satisfaction and quality of surgical care in US hospitals. Ann Surg 2015 Jan;261(1):2-8 [FREE Full text] [doi: 10.1097/SLA.0000000000000765] [Medline: 24887985]
29. Kutney-Lee A, McHugh MD, Sloane DM, Cimiotti JP, Flynn L, Neff DF, et al. Nursing: a key to patient satisfaction. Health Aff (Millwood) 2009 Jan;28(4):669-677 [FREE Full text] [doi: 10.1377/hlthaff.28.4.w669] [Medline: 19525287]

30. Fenton JJ, Jerant AF, Bertakis KD, Franks P. The cost of satisfaction: a national study of patient satisfaction, health care utilization, expenditures, and mortality. Arch Intern Med 2012 Mar 12;172(5):405-411. [doi: 10.1001/archinternmed.2011.1662] [Medline: 22331982]
31. Tevis SE, Kennedy GD, Kent KC. Is there a relationship between patient satisfaction and favorable surgical outcomes? Adv Surg 2015 Sep;49(1):221-233 [FREE Full text] [doi: 10.1016/j.yasu.2015.03.006] [Medline: 26299501]
32. Kahn C, Sandberg C. The next generation of measuring patient experience. Health Affairs. 2019. URL: https://www.healthaffairs.org/do/10.1377/hblog20190307.766083/full/ [accessed 2021-08-05]
33. Halaska C, Sachs P, Sanfilippo K, Lin C. Patient attitudes about viewing their radiology images online: preintervention survey. J Med Internet Res 2019 Jul 18;21(7):e12595 [FREE Full text] [doi: 10.2196/12595] [Medline: 31322124]
34. Ammenwerth E, Schnell-Inderst P, Hoerbst A. The impact of electronic patient portals on patient care: a systematic review of controlled trials. J Med Internet Res 2012 Nov 26;14(6):e162 [FREE Full text] [doi: 10.2196/jmir.2238] [Medline: 23183044]
35. van der Vaart R, Drossaert CH, Taal E, Drossaers-Bakker KW, Vonkeman HE, van de Laar MA. Impact of patient-accessible electronic medical records in rheumatology: use, satisfaction and effects on empowerment among patients. BMC Musculoskelet Disord 2014 Mar 26;15(1):102 [FREE Full text] [doi: 10.1186/1471-2474-15-102] [Medline: 24673997]
36. Tuil WS, Verhaak CM, Braat DD, de Vries Robbé PF, Kremer JA. Empowering patients undergoing in vitro fertilization by providing Internet access to medical data. Fertil Steril 2007 Aug;88(2):361-368. [doi: 10.1016/j.fertnstert.2006.11.197] [Medline: 17416366]
37. Goldberg HI, Ralston JD, Hirsch IB, Hoath JI, Ahmed KI. Using an internet comanagement module to improve the quality of chronic disease care. The Joint Commission Journal on Quality and Safety 2003 Sep;29(9):443-451. [doi: 10.1016/s1549-3741(03)29053-5]
38. Ruggiero TE. Uses and gratifications theory in the 21st century. Mass Commun Soc 2000 Feb;3(1):3-37. [doi: 10.1207/s15327825mcs0301_02]
39. Weaver DH. Communication research in the 1990s: new directions and new agendas. In: Gaunt P, editor. Beyond Agendas: New Directions in Communication Research. Westport, CT: Greenwood Press; 1993:199-220.
40. Cho J, De Zuniga H, Rojas H, Shah D. Beyond access: The digital divide and Internet uses and gratifications. It & Society. 2003. URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.119.5673&rep=rep1&type=pdf [accessed 2021-08-05]
41. Grellhesl M, Punyanunt-Carter NM. Using the uses and gratifications theory to understand gratifications sought through text messaging practices of male and female undergraduate students. Comput Hum Behav 2012 Nov;28(6):2175-2181. [doi: 10.1016/j.chb.2012.06.024]
42. Acquisti A. Privacy in electronic commerce and the economics of immediate gratification. In: Proceedings of the 5th ACM conference on Electronic commerce. 2004 Presented at: EC04: Fifth ACM Conference on Electronic Commerce 2004; May 17 - 20, 2004; New York NY USA p. 21-29. [doi: 10.1145/988772.988777]
43. Carlston D. The Oxford Handbook of Social Cognition. Oxford, United Kingdom: Oxford University Press; 2013:1-948.
44. Isen AM, Daubman KA, Nowicki GP. Positive affect facilitates creative problem solving. J Personal Soc Psychol 1987;52(6):1122-1131. [doi: 10.1037/0022-3514.52.6.1122]
45. Estrada CA, Isen AM, Young MJ. Positive affect improves creative problem solving and influences reported source of practice satisfaction in physicians. Motiv Emot 1994 Dec;18(4):285-299. [doi: 10.1007/bf02856470]
46. Lyubomirsky S, King L, Diener E. The benefits of frequent positive affect: does happiness lead to success? Psychol Bull 2005 Nov;131(6):803-855. [doi: 10.1037/0033-2909.131.6.803] [Medline: 16351326]
47. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425. [doi: 10.2307/30036540]
48. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: The PANAS scales. J Personal Soc Psychol 1988;54(6):1063-1070. [doi: 10.1037/0022-3514.54.6.1063]
49. Bandura A. Social cognitive theory. In: Van Lange P, Kruglanski A, Higgins E, editors. Handbook of Theories of Social Psychology. Thousand Oaks, California, United States: SAGE Publications Ltd; 2012:349-373.
50. Zabihi R. The role of cognitive and affective factors in measures of L2 writing. Writ Commun 2017 Nov 13;35(1):32-57. [doi: 10.1177/0741088317735836]
51. Pérez-Fuentes MD, Linares JJ, Jurado MD, Márquez MD, Martínez Á. The mediating role of cognitive and affective empathy in the relationship of mindfulness with engagement in nursing. BMC Public Health 2020 Jan 07;20(1):16 [FREE Full text] [doi: 10.1186/s12889-019-8129-7] [Medline: 31910847]
52. DeVellis R. Scale Development: Theory and Applications. Thousand Oaks, California, United States: SAGE Publications, Inc; 2016:1-280.
53. Netemeyer R, Bearden W, Sharma S. Scaling Procedures: Issues and Applications. Thousand Oaks, California, United States: Sage Publications Inc; 2003:1-224.
54. Health Information National Trends Survey (HINTS). National Cancer Institute. URL: https://hints.cancer.gov/ [accessed 2021-08-05]
55. Jarvis C, MacKenzie S, Podsakoff P. A critical review of construct indicators and measurement model misspecification in marketing and consumer research. J Consum Res 2003 Sep;30(2):199-218. [doi: 10.1086/376806]

56. Podsakoff PM, MacKenzie SB, Lee J, Podsakoff NP. Common method biases in behavioral research: a critical review of the literature and recommended remedies. J Appl Psychol 2003 Oct;88(5):879-903. [doi: 10.1037/0021-9010.88.5.879] [Medline: 14516251]
57. Gefen D, Straub D. A practical guide to factorial validity using PLS-Graph: tutorial and annotated example. Commun Asso Info Syst 2005;16:5. [doi: 10.17705/1cais.01605]
58. Nunnally J. Psychometric Theory. New York: McGraw-Hill; 1978:1-640.
59. Pavlou PA, El Sawy OA. From IT leveraging competence to competitive advantage in turbulent environments: the case of new product development. Info Syst Res 2006 Sep;17(3):198-227. [doi: 10.1287/isre.1060.0094]
60. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. J Mark Res 2018 Nov 28;18(1):39-50. [doi: 10.1177/002224378101800104]
61. Lowry PB, Gaskin J. Partial Least Squares (PLS) Structural Equation Modeling (SEM) for building and testing behavioral causal theory: when to choose it and how to use it. IEEE Trans Profess Commun 2014 Jun;57(2):123-146. [doi: 10.1109/tpc.2014.2312452]
62. Zajonc RB. Attitudinal effects of mere exposure. J Personal Soc Psychol 1968;9(2, Pt.2):1-27. [doi: 10.1037/h0025848]
63. Zajonc R. Mere exposure: a gateway to the subliminal. Curr Dir Psychol Sci 2016 Jun 23;10(6):224-228. [doi: 10.1111/1467-8721.00154]
64. Zajonc R. Exposure effects. In: Manstead A, Frijda N, Fischer A, editors. Feelings and Emotions: The Amsterdam Symposium. Cambridge: Cambridge University Press; 2004:194-203.
65. Dimmick J, Kline S, Stafford L. The gratification niches of personal e-mail and the telephone. Commun Res 2016 Jun 30;27(2):227-248. [doi: 10.1177/009365000027002005]

Abbreviations

AST: Adaptive Structuration Theory
AVE: average variance extracted
HCAHPS: Hospital Consumer Assessment of Healthcare Providers and Systems
TAM: Technology Acceptance Model
UTAUT: Unified Theory of Acceptance and Use of Technology

Edited by R Kukafka; submitted 03.05.20; peer-reviewed by Anonymous, M Keller, S Bassett, S Bidmon; comments to author 25.05.20; revised version received 15.09.20; accepted 13.07.21; published 27.08.21.

Please cite as:
Kinney AP, Sankaranarayanan B
Effects of Patient Portal Use on Patient Satisfaction: Survey and Partial Least Squares Analysis
J Med Internet Res 2021;23(8):e19820
URL: https://www.jmir.org/2021/8/e19820
doi:10.2196/19820
PMID:34448712

©Aaron P Kinney, Balaji Sankaranarayanan. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Mental Health Care Professionals’ Appraisal of Patients’ Use of Web-Based Access to Their Electronic Health Record: Qualitative Study

Antonius Mattheus van Rijt[1*], MSc, MD; Pauline Hulter[1*], MSc; Anne Marie Weggelaar-Jansen[1,2], MCM, PhD; Kees Ahaus[1], PhD; Bettine Pluut[1], PhD

[1] Erasmus School of Health Policy & Management, Erasmus University Rotterdam, Rotterdam, Netherlands

[2] Eindhoven University of Technology, Eindhoven, Netherlands

[*] these authors contributed equally

Corresponding Author:
Bettine Pluut, PhD
Erasmus School of Health Policy & Management
Erasmus University Rotterdam
Bayle Building - Campus Woudestein
Burgemeester Oudlaan 50
Rotterdam, 3062 PA
Netherlands
Phone: 31 10 408 8555
Email: pluut@eshpm.eur.nl

Abstract

Background: Patients in a range of health care sectors can access their medical health records using a patient portal. In mental health care, the use of patient portals among mental health care professionals remains low. Mental health care professionals are concerned that patient access to electronic health records (EHRs) will negatively affect the patient’s well-being and privacy as well as the professional’s own workload.

Objective: This study aims to provide insights into the appraisal work of mental health care professionals to assess and understand patient access to their EHRs through a patient portal.

Methods: We conducted a qualitative study that included 10 semistructured interviews (n=11) and a focus group (n=10). Participants in both the interviews and the focus group were mental health care professionals from different professional backgrounds and staff employees (eg, team leaders and communication advisors). We collected data on their opinions and experiences with the recently implemented patient portal and their attempts to modify work practices.

Results: Our study provides insights into mental health care professionals’ appraisal work to assess and understand patient access to the EHR through a patient portal. A total of four topics emerged from our data analysis: appraising the effect on the patient-professional relationship, appraising the challenge of sharing and registering delicate information, appraising patient vulnerability, and redefining consultation routines and registration practices.

Conclusions: Mental health care professionals struggle with the effects of web-based patient access and are searching for the best ways to modify their registration and consultation practices. Our participants seem to appraise the effects of web-based patient access individually. Our study signals the lack of systematization and communal appraisal. It also suggests various solutions to the challenges faced by mental health care professionals. To optimize the effects of web-based patient access to EHRs, mental health care professionals need to be involved in the process of developing, implementing, and embedding patient portals.

(J Med Internet Res 2021;23(8):e28045) doi:10.2196/28045

KEYWORDS

patient portals; eHealth; mental health care professionals; mental health; eMental health; mental health care; patient-accessible; electronic health records; Open Notes; normalization process theory; NPT

Introduction

Background

The number of patient portals is increasing rapidly in all health care sectors. Through these patient portals, patients have gained the ability to access their medical health records on the internet. A patient portal is a form of eHealth that can be defined as “provider-tethered applications that allow patients to access, but not to control, certain health care information (eg, their EHR [electronic health record]) and provide communication and administrative functions (eg, secure messaging, appointment booking, and prescription refill requests)” [1]. Research has shown that, in mental health care, the use of a patient portal can have a positive effect on patient activation, recovery, and organizational efficiency [2]. In the same study, mental health care professionals were involved during implementation and were trained to use the patient portal [2]. Furthermore, the relationship between the patient and their mental health care professional can improve, provided the mental health care professional has an open attitude, and the medical record is unique, individualized, and detailed [3]. Another study showed that mental health care professionals could feel uncomfortable because they experience reduced control over the information flow when patients can access their health information on the internet [4]. Overall consequences can be positive, for example, improved registration (ie, *documentation*) and consultations (ie, *visits*) with patients or negative, for example, reduced documentation by mental health care professionals. This suggests that the positive effects of web-based patient access partly depend on the registration practices of the mental health care professional and the ways in which they communicate with their patients [3,4]. Therefore, this study explores the appraisal work carried out by mental health care professionals shortly after the introduction of web-based patient access and sheds light on the challenges mental health care professionals face when trying to make a patient portal work for them and the patient. To gain insight into the challenges of mental health care professionals, we use the normalization process theory (NPT), which helps to understand how new technologies and practices are embedded and integrated into existing work practices [5]. This theory “identifies, *characterises* and explains mechanisms that have been empirically demonstrated to motivate and shape implementation processes and affect their outcomes” [6]. NPT includes a model that explains what health care professionals go through when embedding a new technology which, in this study, we have applied to web-based patient access through a patient portal [5-8]. This paper focuses on one of the key constructs of NPT, reflexive monitoring. Reflexive monitoring concerns the appraisal activities that health care professionals do to assess and understand the ways in which a new set of practices affects them and others around them. For patient portals, the focus is on how patients’ web-based access to sensitive data in the EHR affects mental health care professionals, their patients, and the relationship between them. Reflexive monitoring sheds light on the individual mental health care professionals’ appraisal work shortly after the implementation of web-based patient access. Reflexive monitoring involves four components: (1) *systematization*, which involves collecting information about formal (eg, research results) or informal (eg, anecdotal examples) evidence; (2) during *communal appraisal*, individuals work together to evaluate the worth of, in this instance, patient portals and related working routines; (3) through *individual appraisal*, individuals work experientially to appraise the effects on them and the contexts in which they are set; and (4) *reconfiguration* involves attempts to redefine procedures or modify practices, and perhaps, here, even to change the shape of the patient portal itself, to make the patient portal work [5].

Little is known about the appraisal work of mental health care professionals during the embedding of a patient portal. We do know that patient access to medical health records in mental health care has always been a sensitive subject. In the early 1990s, researchers raised the question of whether reading psychiatric case–related notes could be considered offensive [9]. Especially in mental health care, doctors’ notes often contain sensitive information concerning the mental state of the patient [10]. Research suggests that mental health care professionals think there is a risk that patients disagree with the content of the notes or misinterpret the content, and therefore, patients could be upset [9]. This can cause a patient to become concerned or confused and even to respond angrily. In addition to these specific concerns over sensitive information in mental health care, mental health care professionals share the wider concerns of their colleagues in hospital care [10-12]. In total, 2 studies point to a possible higher work burden caused by increased communication with patients and to a fear of lawsuits or claims for damages [10,13]. However, on the other hand, most mental health care professionals believe that patients will better remember their treatment plans and will be better prepared for appointments [10].

Objectives

This study focuses on appraisal work by mental health care professionals shortly after the implementation of web-based patient access through a patient portal and shows how mental health care professionals try to make sense of this new technology by appraising the effects of the portal and by attempting to modify registrations and consultation practices. Furthermore, our study answers the question of what mental health care professionals do to assess and understand patient access to the EHR through a patient portal.

Methods

Overview

For this qualitative study, 10 interviews with a total of 11 mental health care professionals and, later, a focus group, were conducted in a Dutch mental health care organization. This organization (2100 full-time equivalents) offers mental health care, well-being, and social services for approximately 32,000 inpatients and outpatients of all ages. In January 2019, the organization implemented a patient portal for patients to access their EHRs. All patients were able to read notes, letters, and other information in their EHRs after a period of 30 days. Mental health care professionals cannot determine whether a patient uses web-based access. Medical notes were not accessible by patients if they were marked as a draft, but drafts would eventually have to be marked as final before a course of treatment could be closed. After implementation, a personal notes tab was added for the mental health care professionals. These notes were not visible to colleagues or patients.

Recruitment and Selection

The objective of recruiting study participants was to include mental health care professionals working in diverse focus areas and with different professions within the same mental health care organization. Recruitment, selection, interviews, and focus group were conducted in the spring of 2019. Participants were selected in two ways: by an open invitation on the intranet (n=6) and then through snowballing (n=5). The latter involved asking existing participants if they knew of others who might be willing to be interviewed [14]. All mental health care professionals who expressed willingness to be interviewed were included in the study (Table 1). During the interviews, it became apparent that both supporters and opponents of the patient portal participated in the study.

Participants in the focus group were identified by the head of the computerization and automation department using purposeful sampling (Table 2). This provided a broader range of professions than the interviewee group and included some who had been involved in the implementation of the patient portal.

Table 1. Characteristics of the participants of the interviews.

Participant	Sex	Age (years)	Profession	Focus area
1.1	Female	43	Clinical psychologist	Development disorders
1.2	Female	38	Nurse practitioner	Hospital psychiatry
1.3	Male	51	Nurse practitioner	Anxiety and mood
1.4	Male	54	Nurse practitioner	Elderly
1.5	Female	49	Psychiatrist	Personality disorders
1.6	Female	53	Psychiatrist	Addiction
1.7	Female	54	Psychiatrist	Personality disorders
1.8	Female	60	Psychiatrist	Elderly
1.9	Female	27	Psychologist	Development disorders
1.10	Male	39	Psychologist	First level health care
1.11	Female	61	Psychotherapist and team leader care	Forensic

Table 2. Characteristics of the participants of the focus group.

Participant	Sex	Age (years)	Profession
2.1	Female	53	Functional application manager
2.2	Female	28	Coordinator health care innovation
2.3	Male	57	Team leader anxiety and mood
2.4	Female	52	Team leader specialist diagnosis and treatment
2.5	Male	50	Functional application manager
2.6	Female	38	Team leader anxiety and mood
2.7	Male	Unknown	Computerization and automation
2.8	Male	37	Client council
2.9[a]	Female	27	Psychologist—development disorders
2.10	Female	37	Strategic marketing and communication advisor

[a] Also an interviewee (participant 1.9).

Interviews and Focus Group

Before the interviews and the focus group, participants signed an informed consent form and consented to being audio recorded and the use of the data for research.

One researcher (AMvR) conducted the interviews and the focus group, following a predefined topic list (Multimedia Appendix 1), which was based on earlier research on patient portals [11,15,16]. The topic list for the focus group was also based on the results of the interviews' analysis (Multimedia Appendix 2). During the interviews, participants were asked for their views on and experiences with the potential benefits and risks of patient access to the EHR through a patient portal and possible solutions to reduce the identified risks. The interviews lasted 50 minutes on average (range 30-73 minutes). Most interviews took place in a face-to-face setting. Only one interview was conducted on the internet through Skype because of the geographically distant location of the participant [17].

The focus group was intended to check and enrich the results of the interviews while creating room for elaboration [18]. Participants in the focus group were presented a tentative analysis of the interviews, after which discussion took place according to the predefined topic list (Multimedia Appendix 2). This led to in-depth discussions on the perspectives of mental health care professionals on patient access to their EHR [19]. The focus group lasted 75 minutes and was conducted in a face-to-face setting.

Analysis

The interviews and the focus group were audio recorded and then transcribed verbatim. First, we followed an inductive approach to analyze the data, in which we repeatedly examined which themes emerged from our data [20]. Second, we took a deductive approach, in which we looked at our data through the lens of NPT to analyze the different components of reflexive monitoring by mental health care professionals. Combined, our analysis can be described as abductive [21]. We coded the data in three steps: open, axial, and selective [22]. Keywords were coupled to certain fragments of the transcripts (Multimedia Appendix 3). Using these keywords, connections were made between different fragments of various transcripts. Thereafter, these keywords were regrouped and formed the basis for drawing conclusions from this research. All the interviews were first individually and separately coded by 2 members of the research team (AMvR and BP), after which these codes and themes were discussed, reviewed, and adjusted if necessary until a consensus was reached (AMvR and BP). Subsequently, we discussed and adjusted the outcomes where necessary with the other members of the research team [20]. The analysis was computer-assisted using ATLAS.ti software (version 8; Scientific Software Development GmbH) [23].

Results

Overview

The aim of our study was to provide insights into the appraisal work that mental health care professionals do to assess and understand patient access to the EHR through a patient portal. A total of four interrelated topics emerged from the data analysis: (1) appraising the effect on the patient-professional relationship, (2) appraising the challenge of sharing and registering sensitive information, (3) appraising patient vulnerability, and (4) redefining consultation routines and registration practices.

Our analysis showed that there were both opponents and supporters of web-based patient access among our participants. The following two quotes illustrate the strong differences in opinions among the interviewed mental health care professionals:

> *I must honestly say that I have not thought about the possible benefits. I only saw disadvantages, felt that I have to be very careful. That was my first response.* [P 1.8, psychiatrist]
>
> *I think it is a greater risk if patients do not have online access.* [P 1.5, psychiatrist]

Furthermore, our analysis showed that opponents tend to focus on their concerns and have difficulty mentioning the benefits of web-based patient access. When mentioning an advantage, they sometimes immediately denounce the advantages. For example, when asked about the benefit of web-based patient access, one opponent answered:

> *I might forget to write something down, patients can mention this. So that could be an advantage. However, I must say now that I mention it, I am also immediately afraid that this will cause a lot of extra work.* [P 1.8, psychiatrist]

Appraising the Effect on the Patient-Professional Relationship

One of the effects our participants perceived with patient access to the EHR through a patient portal is that it changes the patient-professional relationship.

The first way in which the patient-professional relationship could be changed by patient access is through feedback provided by patients on the content of the EHR. Participants explained that when patients believe the information they read is incorrect or that information is missing, this can be adjusted, leading to therapeutic gain and a new kind of conversation between the patient and professional. One participant illustrated this as follows:

> *If it [patient access to the EHR] produces complaints, you have to do something about it. If people are correct, they are right to complain and you should not be uncooperative but adjust something. And, it is possible that if you can talk about it with a patient, this could improve the therapeutic relationship.* [P 1.4, nurse practitioner]

However, participants also mentioned that these extra questions, comments, or even complaints from patients, take time to answer, and mental health care professionals might need to change their records afterwards.

> *I am a little bit afraid that the people that will be looking [in their medical record], are the people that will have a lot of criticism on what I have written. They will say I did not mean this, I meant it like this.* [P 1.8, psychiatrist]

Second, our participants argued that patients being able to read their EHR both before and after a consultation with the mental health care professional could enable them to be better prepared for their appointments, and therefore enhance the quality of the conversation between patients and professionals.

Third, participants argued that patients who read the information in their EHR could be more aware of their treatment and feel more like an equal to the mental health care professional. Mental health care professionals could also help create a sense of shared responsibility for the treatment by encouraging patients to study their health information in the patient portal. One participant illustrated the following:

> *Very often I hear: "Oh, I do not know where my treatment plan is." You can point it out and mention*

> *that it is something that belongs to both of us.* [P 1.7, psychiatrist]

As the examples above illustrate, our participants believed that web-based patient access could: (1) increase the therapeutic gain, (2) improve the patient's preparation for a consultation, and (3) improve the involvement of patients in their treatment. However, participants also feared that web-based patient access might cost a lot of valuable time and that patients' reading notes could have a negative effect on the patient-professional relationship, which is further described in the next section.

Appraising the Challenge of Sharing and Registering Sensitive Information

Participants admitted that they were struggling with the way they formulated information for the EHR. Medical information in mental health care is often subjective, and writing down a diagnosis is a delicate balance, which is illustrated by one participant's reflection:

> *Especially if one [health professional] did not consider it [web-based patient access], one could have written in a somewhat unsophisticated way in the medical record: "This is typical of borderline behavior.," while not seeing*
>
> *the patient as borderline.* [P 1.7, psychiatrist]

Some participants were worried that patients might feel insulted, misinterpret the information given, or feel unheard when reading the information in the EHR, which could reduce trust in the treatment or even withdrawal from the care program:

> *[...] people who are attached in an unsafe way will very quickly feel let down, and that is also possible through text, which, getting back to the therapeutic relationship, can of course deteriorate, and that would be a pity.* [P 1.7, psychiatrist]

On the other hand, participants mentioned that such information is an important part of the psychiatric examination and might be important for colleagues to know. If some information is not appropriate for patients to read in their EHR, then mental health care professionals can be reluctant to write it down. One participant illustrated the following:

> *Let's assume I see someone who looks dirty or with poor hygiene, then I have a hard time writing that down.* [P 1.7, psychiatrist]

Besides being subjective, information on mental health care is also often sensitive. Participants argued that patients might become overwhelmed and eventually relapse (a deterioration in the mental health of an individual who was controlling their mental illness) because of the amount or content of the information they have at their disposal with access to their EHR. One participant said:

> *[...] there are people who can go backwards over small details, such as "I did not study for seven months but eight" [...]* [P 1.1, clinical psychologist]

This view was confirmed by the participants in the focus group, where a team leader mentioned that he observed that his colleagues were less detailed in their registration:

> *[...] you also hear that care providers are more aware of what they write in their report, and therefore are more factual and less informative [...]* [P 2.3, team leader anxiety and mood]

In addition, our participants explained that an mental health care professionals' report of a consultation might reveal that the patient and the professional had experienced their conversation quite differently and felt differently about what was most important or would therefore summarize the highlights and conclusions differently. Our participants stated that this is not unusual with mental health care and occurs less with physical issues. One participant explained the following:

> *I wrote a note in the medical record in a certain way, but maybe the other person (the patient) experienced a different conversation.* [P 1.7, psychiatrist]

Our participants had various views on entering information that is not yet intended for patients. For instance, collateral history might contain sensitive and possibly offensive information and may not always be suitable for patients or known by them. Our participants experience this ethical dilemma: they are not sure whether they should write down sensitive information and whether this information belongs to the patient's EHR. One participant saw it as a moral dilemma whether to enter certain information or not because it could be beneficial to the treatment but also involves the risk of harming the patient. According to our participants, some information would not be beneficial for patients if they saw it. One participant offered the following example:

> *I have had a patient, [...] that girl was sixteen years old and her mother was pregnant through the daughter's boyfriend, and that was written in her medical record, [...] but the girl did not know. [...] It was relevant to the background about the girl's tangled family situation where all kinds of things had occurred, with very unusual relationships.* [P 1.4, nurse practitioner]

Another example of doubts about entering information that is not yet, if ever, intended for patients is over certain treatment plans, with participants worrying that they might no longer work if patients can read about them. One participant illustrated a situation where a patient's husband and her general practitioner thought her situation was deteriorating, but the patient herself did not agree and did not want any kind of treatment. The participant called the patient, and the patient made clear that she did not want any treatment. The participant said the following:

> *I will make a note of that: "spoken today, clearly different than yesterday, much angrier today, does not want an appointment, does extensively talk about it, agreed that I will call her again next week to see if there are any possibilities then, otherwise I will ask her husband to come here with her," that is my plan. I did not tell her all of it [...]* [P 1.3, nurse practitioner]

The issue over treatment plans led to mental health care professionals doubting whether patients should have real-time access to their EHR rather than a 30-day delay. When patients need acute care or are compelled to receive care, for example, in crisis situations, real-time patient access might lead to dangerous situations if patients read what mental health care professionals are planning. One participant stated as follows:

> *It is possible that when he [the patient] reads this and thinks: "Hey, they are on my doorstep tomorrow [for an involuntary admission], you know what, I will end it [his life] before they arrive."* [P 1.9, psychologist]

Despite the dangers of disclosing information to patients, our participants were aware that not entering their thoughts in the EHR also carried risks. Information might otherwise be lost or colleagues are no longer fully informed about certain patients. In crisis situations, where mental health care professionals work in shifts, the peer transfer of information is seen as important by our participants. Furthermore, a participant in the focus group mentioned that mental health care professionals are responsible for what they enter, but also if they fail to enter information that might be of importance later:

> *[...] suppose you have seen or recognized something, and you did not want to write it down for whatever reason, but it does have an influence on a future course of the treatment, or possibly a crisis situation, and you say: "well, I did see or spot that earlier on," you are responsible for that.* [P 2.4, team leader specialist diagnosis and treatment]

This influences the way our participants work individually and together, especially when they disagree about certain issues and have yet to make decisions about how they redefine their registration and consultation practices.

In summary, when sharing and registering delicate information, our participants struggle individually with the way they should write information in the EHR and are afraid that (1) it could reduce patients' trust in their treatment because patients misinterpret the information they have access to, (2) mental health care professionals might enter information that is inappropriate for patients to read, (3) patients might become overwhelmed by the amount or content of the entered information, (4) it might show to patients that professionals have experienced their conversation quite differently than they did themselves, and (5) there is no place to write down information that is not yet, if ever, intended for the patient to read. This shows that our participants, as individuals, have thought deeply about how to make mental health care patients' access to their EHR work. There were disagreements over entering information that was not yet, if ever, intended for patients. Whether or not to enter certain information seemed to be a moral dilemma because it could be beneficial to the treatment but also involves a risk of harm to the patient.

Appraising Patient Vulnerability

Our participants worried that patients could become more vulnerable with web-based access to their EHRs. They were concerned that they had little control over how patients would act on this information in the EHR and are also afraid that patients might, for example, deteriorate after reading their own medical record.

As mentioned in the previous paragraph, information on mental health care is often sensitive. Participants were afraid that this could overwhelm patients and possibly cause a harmful relapse:

> *[...] during meetings we discuss whether an admission to the ward would be an option. If you write down that you consider this, he [the patient] might get upset or deteriorate. [...] The same goes for our considerations, should we write down something else to prevent a patient from deteriorating?* [P 1.9, psychologist]

When asked what is meant by deterioration, a participant answered as follows:

> *[...] a patient getting completely disordered, mentally stuck, upset, a breach of trust with their mental health care professionals [...].* [P 1.1, clinical psychologist]

Participants explained that patients could easily print or download their own medical records, after which they could share this with *inappropriate* people. In this way, sensitive information may fall into the wrong hands. A third party, such as a curious spouse, could also gain access to the EHR for wrong reasons. Especially in mental health care, patients are often vulnerable and easily influenced by relatives. One participant stated as follows:

> *A disadvantage could be that someone else gets access to the password or login codes, that could of course be a risk. With certain treatments, you do not want a partner to know certain things, [...] however they [relatives] can be persuasive and demand access from a patient.* [P 1.10, psychologist]

Our participants were unsure who would be responsible for the potentially reckless handling of information from the EHR by the patient. They also doubted whether it would be sufficient if mental health care professionals warn patients about the sensitive nature of the information. One participant, however, stated that sharing health data was the responsibility of the patients:

> *The patient has access, so I think it is their responsibility. I think the content and the correctness of the content is the responsibility of the health professional.* [P 1.10, psychologist]

Another participant mentioned an extreme example of what could happen when patients share their own medical information, for example, to show that they are discontent with their treatment, but emphasized that this is the patient's own responsibility:

> *If the patient thinks: "I will go to Story or RTL boulevard [national media] with my medical record, which sometimes happens, then they can do it."* [P 1.11, psychotherapist and team leader care]

In summary, our participants were afraid that they had much less control over what patients do with the information in the EHR and wonder who is responsible for sharing information. It seems that patient access raises many uncertainties concerning individual personal relationships with a patient.

Redefining Consultation Routines and Registration Practices

Reflecting on the effects and struggles of entering sensitive information in an EHR, our participants suggested various solutions in terms of modifying their registration practices. However, those who opposed the idea of web-based patient access were not convinced that those solutions would really work, as they often also mentioned the possible disadvantages of the suggested solution.

Solution 1: Draft Notes for Colleagues

The first solution suggested by the participants was to write draft notes for colleagues. This is a temporary solution, in that draft notes will not be immediately visible in the patient portal but will need to be marked as final, and hence become visible, before a treatment can be closed.

Solution 2: Making Personal Notes Visible for Colleagues

The second solution was to make the personal notes tab visible to colleagues. Although this prevents the loss of access to information in, for example, crisis situations, this also reduces the transparency of information for patients because a hidden *shadow file* is created.

Solution 3: Discussing Information With Patients Before Registration

A third suggested solution was to discuss information with patients before the mental health care professionals register this information. In this way, they can ensure that there is no new information in the EHR should the patient choose to access it. A participant in the focus group explained this as follows:

> *[...] it is quite difficult in that you cannot write down your considerations, but I think it is also a stimulant to share your considerations with the patient a lot more, by which you give a patient more space and influence, which causes the treatment relationship to become more equal [...].* [P 2.6, team leader anxiety and mood]

However, participants acknowledged that this third option was only workable if they discuss the information directly during a consultation. If not, if mental health care professionals delay registration, this increases the risk of mistakes and lost information because of memory shortfalls. At the same time, our participants commented that they often let a conversation sink in and write the report later:

> *[...] of course it remains difficult, when you walk back into your room and you smell alcohol [lingering from the patient] after the end of a consultation. Where do you record this, as you have not yet discussed it with the patient, but it is important information, these are difficult things.* [P 2.6, team leader anxiety and mood]

Solution 4: Registering Information Together With the Patient

A fourth solution that is mostly mentioned by supporters is the practice of registering information together with the patient. This collaborative practice could even become a form of treatment. Our participants felt that it depends on the patient whether this would be a workable solution, and two possible obstacles were raised by opponents. First, it was noted by the participants that certain patients (eg, psychotic patients or those with developmental disorders) are not capable of writing notes along with their mental health care professional. For example, an opponent mentioned that patients with a developmental disorder are often overstimulated after a consultation and would not be able to contribute to writing notes:

> *The argument is: “you have to write [in the medical record] together with your patient, use the last ten minutes of your consultation.” However, that does not work with our patients. They are completely overstimulated after half an hour, they cannot immediately reflect on what happened.* [P 1.1, clinical psychologist]

However, when a supporter was confronted with this concern, she responded as follows:

> *It can be an extra effort, but that is also part of the dynamics of that treatment. [...] No psychiatrist is made to treat everyone, [...], so I guess choose your patient population according to that.* [P 1.5, psychiatrist]

Second, some participants were concerned that writing notes with the patient would eat into the already limited time for consultation.

Solution 5: Introducing Patients to Web-Based Access at the Beginning of Treatment

The final solution was to introduce patients to web-based access to their EHRs at the beginning of their treatment. Mental health care professionals could then explain the risks and benefits of web-based patient access and decide together with the patient whether the patient would use it. One participant said the following:

> *Sometimes the risks have to be pointed out to a patient, as I just said, you can send a copy of your letter but watch out when you use it in court, so they need to be informed about the risks.* [P 1.6, psychiatrist]

Our participants would like more support on what kind of information patients can read in the EHR and how they should write sensitive information in the EHR. This could provide them with more knowledge and enable them to experiment with web-based patient access and to evaluate the outcomes together. Our participants said it was unclear to them what kind of information patients could read through the patient portal and on which terms. Furthermore, participants commented that they had only limited experience with the patient portal because it had only just been implemented. Indeed, most participants had no personal experience with patients accessing their EHR at all. However, some participants were able to report on one encounter with a patient who had read their EHR and then regretted doing so:

> *Some patients get overwhelmed by the amount of information, one patient said the following: “I just*

> *regret looking because I started and I got so much information, well, I got really upset, then I stopped.”* [P 1.7, psychiatrist]

As illustrated earlier, our participants were individually able to come up with five solutions that they believed could make patient access to the EHR work for them as well as for their patients. However, it would appear that our participants needed more support on how the portal works so that they could actually experiment with their ideas on working with web-based patient access and evaluate these experiments.

Discussion

Principal Findings

This study seeks to provide insights into the appraisal work that mental health care professionals do to assess and understand patient access to their EHRs through a patient portal. By interviewing 11 mental health care professionals and conducting a focus group discussion, we learned that mental health care professionals struggle with how to weigh up the potential benefits and risks they perceive and are trying to work out what they can do themselves to make the portal work for them and for their relationship with their patients.

Our results show that mental health care professionals struggle with various aspects of patient access to the EHR and with entering what they perceive as sensitive information into the EHR. First, we looked at the ways in which mental health care professionals appraise the effect of web-based patient access on their relationship with the patient. Second, we report how mental health care professionals fear that some patients are too vulnerable to handle the new possibility of accessing their medical records. Third, we showed the ways in which mental health care professionals address the challenge of registering and discussing delicate information. Finally, we showed how mental health care professionals individually experiment by redefining consultation routines and registration practices.

Our results show that participants are actively engaged in the NPT terms reflexive monitoring, especially the components related to individual appraisal and reconfiguration [5]. Our participants individually appraised the effects of patient access to the EHR (eg, that mental health care professionals should perhaps no longer write so freely in the medical record) and thought about solutions to modify and redefine their registration and consultation practices. Participants mentioned that notes might become less accurate and less detailed to avoid potential harm to the patient, a concern also expressed elsewhere in the literature [10,13,24]. Although some studies show that, in practice, only very few patients are actually harmed [12,25], another study showed that patients could be surprised or hurt when they read information in the medical record that is incorrect, outdated, or new to them [3]. Such patients are then afraid that this incorrect or outdated information might have a negative impact on their treatment if, for example, other mental health care professionals read and act on this information [3]. Other patients commented that this makes them doubt whether their mental health care treatment is useful [3]. The other two NPT components, systematization and communal appraisal, did not appear to take place. As long as mental health care professionals struggle to engage with these two components of reflexive monitoring, embedding web-based patient access in the work practices of mental health care professionals will be hindered. Consequently, we hope that future research will explore the ways in which systematization and communal appraisal can be stimulated during the implementation of web-based patient access in mental health care. In addition, future research could focus on ways to involve opponents of web-based patient access in the process of communal appraisal and reconfiguration.

Furthermore, our results show that participants worry that certain treatment plans and strategies might no longer work if patients can read them. This is a new concern that has not been mentioned in the literature before and is especially relevant as information in the EHR becomes accessible in real time. However, a study on real-time access through a patient portal in hospital care concluded that the limited negative consequences could be mitigated by instruction, education, and preparation of patients by the mental health care professionals [26]. Further research on this topic in mental health care is recommended and could focus on the cocreation of further development of web-based patient access with patients [27].

NPT suggests that appraisal work needs to include communal appraisal if a technology is to become normalized, that is, for it to become an integrated aspect of the mental health care professionals’ work routines. During the interviews, participants suggested various solutions to the struggles they experience with patients having web-based access to their medical health records. Individual mental health care professionals suggesting adaptions to the new service, so that it becomes a normalized practice, is in accordance with the reflexive monitoring component of NPT [5]. Mental health care professionals and the organization as a whole could work on these solutions to eventually embed web-based patient access in their daily work routines. For this to occur, mental health care professionals can discuss their concerns and struggles and cocreate solutions, such as the concerns and solutions expressed and suggested during the interviews [28]. The solutions mentioned in this paper could serve as a starting point but still need to be evaluated in practice.

Limitations

Our study has four limitations. First, our study focused on a specific organization in mental health care with the mental health care professionals involved all having a similar amount of experience with web-based patient access. Furthermore, not all focus areas within mental health care were represented. Therefore, some mental health care divisions, such as forensic psychiatry and primary mental health care, were probably underrepresented. Second, because all the participants actively responded to an open invitation to participate, there is a risk of selection bias. There is also a possibility that only early adopters of the patient portal participated in the interviews, given that the organization implemented the portal in January 2019, and the interviews were conducted in the spring of that year. It might be possible that the participants of this study were not representative of the population of mental health care professionals. However, the interview transcripts show that both proponents and opponents and some mental health care professionals with more neutral views took part in the study. Moreover, it is important to note that, given the very limited time between the implementation of the patient portal and our interviews, most of the worries expressed by the participants were not based on specific personal experiences with web-based patient access. It would be interesting to repeat this study to see whether the mental health care professionals have changed their minds or have experienced the struggles they expected and whether collective experience or evaluations had already occurred. A third limitation is that, apart from one participant in the focus group, the patient perspective was excluded. Further research is needed to explore how the doubts expressed in this study are experienced by patients. Finally, in the topic list, we choose not to explicitly ask participants to reflect on the four different components of reflexive monitoring according to NPT. In contrast, we chose to center the appraisal activities as articulated by the participants themselves. Future research is needed to validate our finding that systematization and communal appraisal are not the predominant components of reflexive monitoring by mental health care professionals.

Comparison With Prior Work

Research shows that patient access through patient portal empowers patients, meaning that patients feel more in control of their mental health care [2,12,29]. A pilot study involving 52 psychiatric patients gaining web-based access to their medical health record found that 82% of the included patients felt more in control of their own treatment because of the possibility of reading their treatment plans and medical notes and knowing what they could expect in their care process [12]. However, our results show that doubts remain as to whether mental health care patients can handle access to their own EHR. For example, our participants were afraid that patients might share their medical records with an unauthorized person or authority, which could make patients more vulnerable to people or institutions with conflicting interests. A recent review similarly raised this concern regarding patients autonomously handling medical information [30]. Another study found that a major barrier to redefining work practices of health care professionals through the use of patient portals in hospital care concerned privacy and security [31]. These examples support our finding that mental health care professionals are struggling to assess and understand the effect of web-based patient access for their patients and their work practices. Further research should confirm our findings and should look for more solutions to reduce the privacy and security concerns of mental health care professionals.

There is a moral dilemma if the benefits of web-based patient access are associated with an increase in patient vulnerability. This has its roots in the normative question of what is *good*. Is it *good* to aim for the benefits of web-based access and increasing empowerment, but possibly also resulting in an increase in patient vulnerability, or is it *good* to prevent an increase in vulnerability that involves withholding possible benefits? And, maybe even more importantly, whose decision is this to make? There are no universal answers to these normative questions, but it is important to recognize and discuss these dilemmas. The thin line between patient autonomy, patient empowerment, and patient vulnerability has been discussed in various studies on patient-centered care, as is evident from a discourse analysis on patient-centeredness, which indeed highlights that there are different views on what is *good* patient care [32]. Some consider patient-centeredness to be a process of *empowering patients*, implying that they believe patients should be given the possibility to view their medical data on the web. Withholding web-based access to medical information for vulnerable patients could be considered unethical in this discourse. Risks are recognized, but empowerment also helps patients to appropriately deal with the risks of web-based patient access. In another discourse, which we label *caring for patients*, people have a more paternalistic view of patient-centeredness and believe that health care professionals should protect patients from risks. Our results indicate that some mental health care professionals doubt that it is their task to protect patients from certain vulnerabilities. However, our participants also commented that not all patients are the same and that patients require tailored care. This reflects the *being responsive* discourse, which argues that patient-centered care is about meeting the specific and highly differing needs of patients. Individual mental health care professionals and organizations as a whole need to determine what patient-centered care means to them and how they want to deal with the moral dilemmas associated with patient access to the EHR. Communal appraisal can be arranged by organizing a moral deliberation, one of the ways to organize a dialog about the moral dilemmas of patient autonomy versus patient vulnerability [33].

As our results indicate, the protection of vulnerable patients might not only be the responsibility of mental health care professionals through individual appraisals. The literature shows that this can also be achieved through laws and regulations [34]. Patients gaining more control over their own EHR falls under the term *informational self-determination*, which is defined as "the ability of a person to determine, in principle, to what extent personal data is used and further disclosed, in view of a self-determined life" [34]. With an increase in informational self-determination, the risk of spreading medical information to parties who are not entitled to it increases. A possible solution could be to implement patient confidentiality, in which medical information managed by patients is legally protected [34]. Further research could explore the feasibility of this concept and look at ways to include mental health care professionals and modify their practices.

Research investigating patient access through a patient portal in hospital care has shown that patients' interests and abilities in using a patient portal are influenced by various factors, including age, health literacy, and level of education [35]. Patients are more likely to use a patient portal if it suits their information needs and has the functionalities they require [35]. Our results show that a possible solution could be to introduce every new patient to web-based patient access with mental health care professionals, discussing with them the possibilities and the possible risks regarding privacy and their responsibilities. This would involve mental health care professionals in (1) collecting information in various ways, such as asking the opinions of patients and colleagues; (2) jointly evaluating how introducing new patients to web-based patient access would work; (3) individually experiencing if introducing every new patient to web-based patient access adds value; and (4) appraising, alone or with each other, if this way of working requires a redefinition of their registration and consultation practices, or even a change in the patient portal itself. In a recent study, some mental health care professionals believed that informing patients about the benefits and risks of reading medical notes was worthwhile [36]. However, in the same study, there were also mental health care professionals who were reluctant to inform patients about this because they feared negative outcomes. The study concluded that clear patient-professional communication about web-based access to medical information would prevent potential harm. However, another study concluded that introducing every patient to web-based access at the beginning of their treatment would be time consuming and might not be feasible [10]. Another option would be a web-based educational program for mental health patients to introduce them to web-based access. Indeed, one study argued that this may help empower patients and increase their active participation in their own care [37]. Another study found that a web-based course for mental health care professionals on web-based patient access in mental health care resulted in a reduction in mental health care professionals' worries about web-based patient access and an improvement in aspects of patient-professional communication [37]. Further research is needed to explore the feasibility of these solutions as a way to modify the practices of mental health care professionals; researchers should also be open to other possible solutions, such as action research, because this can directly improve the embedding of patient EHR access because improvements can be made during the study [38].

Conclusions

This study provides insights into the appraisal work that mental health care professionals do to assess and understand patient access to their EHRs through a patient portal. Our study explores and describes the effects and struggles that mental health care professionals experience with patients having access to their EHR and how they individually experiment to redefine and modify their work practices. One new insight, not previously reported, is that mental health care professionals are concerned that their treatment plans might no longer be effective. In certain situations, such as when patients need acute care or are compelled to receive care, real-time patient access might lead to dangerous situations because patients act before mental health care professionals can carry out their treatment plan. Furthermore, our study signals a lack of systematization and communal appraisal. Our participants predominantly seem to individually appraise the effects of web-based patient access and how they can modify their registration and consultation practices. Future research is needed to investigate the ways in which systematization and communal appraisal can be stimulated.

In addition, future research could investigate the viability of the modifications in consultation routines and registration practices proposed by our participants. Finally, future research could focus on ways to involve opponents of web-based patient access in communal appraisal. The findings of this study can help researchers, project leaders, project staff, policy officers, and mental health care professionals to understand the process of embedding a new technology and the need for communal appraisal. To further improve working with web-based patient access, mental health care professionals need to be involved in evaluations and the further development of patient portals.

Acknowledgments

The authors would like to thank the mental health care organization involved and the individual participants for making this study possible.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Topic list for the interviews.
[DOCX File , 17 KB - jmir_v23i8e28045_app1.docx]

Multimedia Appendix 2
Topic list for the focus group.
[DOCX File , 19 KB - jmir_v23i8e28045_app2.docx]

Multimedia Appendix 3
Analysis of keywords.
[DOCX File , 15 KB - jmir_v23i8e28045_app3.docx]

References

1. Grünloh C, Myreteg G, Cajander A, Rexhepi H. "Why Do They Need to Check Me?" Patient Participation Through eHealth and the Doctor-Patient Relationship: Qualitative Study. J Med Internet Res 2018 Jan 15;20(1):e11 [FREE Full text] [doi: 10.2196/jmir.8444] [Medline: 29335237]
2. Kipping S, Stuckey MI, Hernandez A, Nguyen T, Riahi S. A web-based patient portal for mental health care: benefits evaluation. J Med Internet Res 2016 Nov 16;18(11):e294 [FREE Full text] [doi: 10.2196/jmir.6483] [Medline: 27852556]
3. Cromer R, Denneson LM, Pisciotta M, Williams H, Woods S, Dobscha SK. Trust in mental health clinicians among patients who access clinical notes online. Psychiatr Serv 2017 May 01;68(5):520-523 [FREE Full text] [doi: 10.1176/appi.ps.201600168] [Medline: 28142383]
4. Denneson LM, Pisciotta M, Hooker ER, Trevino A, Dobscha SK. Impacts of a web-based educational program for veterans who read their mental health notes online. J Am Med Inform Assoc 2019 Jan 01;26(1):3-8 [FREE Full text] [doi: 10.1093/jamia/ocy134] [Medline: 30445648]
5. May C, Finch T. Implementing, embedding, and integrating practices: an outline of normalization process theory. Sociology 2009 Jun 15;43(3):535-554. [doi: 10.1177/0038038509103208]
6. May CR, Cummings A, Girling M, Bracher M, Mair FS, May CM, et al. Using normalization process theory in feasibility studies and process evaluations of complex healthcare interventions: a systematic review. Implement Sci 2018 Dec 07;13(1):80 [FREE Full text] [doi: 10.1186/s13012-018-0758-1] [Medline: 29879986]
7. May C. Towards a general theory of implementation. Implementation Sci 2013 Feb 13;8(1):1-14. [doi: 10.1186/1748-5908-8-18]
8. McEvoy R, Ballini L, Maltoni S, O'Donnell CA, Mair FS, Macfarlane A. A qualitative systematic review of studies using the normalization process theory to research implementation processes. Implement Sci 2014 Jan 02;9:2 [FREE Full text] [doi: 10.1186/1748-5908-9-2] [Medline: 24383661]
9. Crichton P, Douzenis A, Leggatt C, Hughes T, Lewis S. Are psychiatric case-notes offensive? Psychiatr Bull R Coll Psychiatr 1992 Nov;16(11):675-677. [doi: 10.1192/pb.16.11.675] [Medline: 11659954]
10. Dobscha SK, Denneson LM, Jacobson LE, Williams HB, Cromer R, Woods S. VA mental health clinician experiences and attitudes toward OpenNotes. Gen Hosp Psychiatry 2016;38:89-93. [doi: 10.1016/j.genhosppsych.2015.08.001] [Medline: 26380876]
11. Kahn MW, Bell SK, Walker J, Delbanco T. A piece of my mind. Let's show patients their mental health records. J Am Med Assoc 2014 Apr 02;311(13):1291-1292. [doi: 10.1001/jama.2014.1824] [Medline: 24691603]
12. Peck P, Torous J, Shanahan M, Fossa A, Greenberg W. Patient access to electronic psychiatric records: A pilot study. Health Policy and Technology 2017 Sep;6(3):309-315. [doi: 10.1016/j.hlpt.2017.06.003]
13. Petersson L, Erlingsdóttir G. Open notes in Swedish psychiatric care (part 1): survey among psychiatric care professionals. JMIR Ment Health 2018 Feb 02;5(1):e11 [FREE Full text] [doi: 10.2196/mental.9140] [Medline: 29396386]
14. Goodman LA. Snowball sampling. Ann Math Statist 1961 Mar;32(1):148-170. [doi: 10.1214/aoms/1177705148]
15. Otte-Trojel T, de Bont A, Rundall TG, van de Klundert J. What do we know about developing patient portals? a systematic literature review. J Am Med Inform Assoc 2016 Apr;23(e1):e162-e168 [FREE Full text] [doi: 10.1093/jamia/ocv114] [Medline: 26335985]
16. Klein JW, Jackson SL, Bell SK, Anselmo MK, Walker J, Delbanco T, et al. Your patient is now reading your note: opportunities, problems, and prospects. Am J Med 2016 Oct;129(10):1018-1021. [doi: 10.1016/j.amjmed.2016.05.015] [Medline: 27288854]
17. Morgan D. Focus group interviewing. In: Handbook of Interview Research: Context and Method. London, UK: SAGE Publications; 2001.
18. Stewart D, Shamdasani P. Focus Groups: Theory and Practice. Newbury Park: Sage; 1991.
19. Liamputtong P. Focus Group Methodology: Principles and Practice. London: Sage Publications Ltd; 2011.
20. Gioia DA, Corley KG, Hamilton AL. Seeking qualitative rigor in inductive research. Organizational Research Methods 2013 Jan;16(1):15-31. [doi: 10.1177/1094428112452151]
21. Gehman J, Glaser VL, Eisenhardt KM, Gioia D, Langley A, Corley KG. Finding theory–method fit: a comparison of three qualitative approaches to theory building. J Manag Inquiry 2017 May 30;27(3):284-300. [doi: 10.1177/1056492617706029]
22. Strauss A, Corbin J. Basics of Qualitative Research: Grounded Theory Procedures and Techniques. Newbury Park: Sage; 1990.
23. ATLAS.ti 8. Windows. Berlin: ATLAS.ti Scientific Software Development GmbH; 2019. URL: https://atlasti.com/ [accessed 2021-08-09]
24. Erlingsdóttir G, Petersson L, Jonnergård K. A theoretical twist on the transparency of open notes: qualitative analysis of health care professionals' free-text answers. J Med Internet Res 2019 Sep 25;21(9):e14347. [doi: 10.2196/14347] [Medline: 31573905]
25. Petersson L, Erlingsdóttir G. Open notes in Swedish psychiatric care (part 2): survey among psychiatric care professionals. JMIR Ment Health 2018 Jun 21;5(2):e10521 [FREE Full text] [doi: 10.2196/10521] [Medline: 29929946]
26. van Kuppenveld SI, van Os-Medendorp H, Tiemessen NA, van Delden JJ. Real-time access to electronic health record via a patient portal: is it harmful? A retrospective observational study. J Med Internet Res 2020 Feb 06;22(2):e13622 [FREE Full text] [doi: 10.2196/13622] [Medline: 32044753]

27. Kildea J, Battista J, Cabral B, Hendren L, Herrera D, Hijal T, et al. Design and development of a person-centered patient portal using participatory stakeholder co-design. J Med Internet Res 2019 Feb 11;21(2):e11371 [FREE Full text] [doi: 10.2196/11371] [Medline: 30741643]
28. van Gemert-Pijnen JE, Nijland N, van LM, 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;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
29. Stein EJ, Furedy RL, Simonton MJ, Neuffer CH. Patient access to medical records on a psychiatric inpatient unit. Am J Psychiatry 1979 Mar;136(3):327-329. [doi: 10.1176/ajp.136.3.327] [Medline: 420331]
30. Laukka E, Huhtakangas M, Heponiemi T, Kujala S, Kaihlanen A, Gluschkoff K, et al. Health care professionals' experiences of patient-professional communication over patient portals: systematic review of qualitative studies. J Med Internet Res 2020 Dec 08;22(12):e21623 [FREE Full text] [doi: 10.2196/21623] [Medline: 33289674]
31. Kooij L, Groen WG, van HW. Barriers and facilitators affecting patient portal implementation from an organizational perspective: qualitative study. J Med Internet Res 2018 May 11;20(5):e183 [FREE Full text] [doi: 10.2196/jmir.8989] [Medline: 29752253]
32. Pluut B. Differences that matter: developing critical insights into discourses of patient-centeredness. Med Health Care Philos 2016 Dec;19(4):501-515 [FREE Full text] [doi: 10.1007/s11019-016-9712-7] [Medline: 27251048]
33. Gill SD, Fuscaldo G, Page RS. Patient-centred care through a broader lens: supporting patient autonomy alongside moral deliberation. Emerg Med Australas 2019 Aug;31(4):680-682. [doi: 10.1111/1742-6723.13287] [Medline: 30924236]
34. Hooghiemstra T. Informationele Zelfbeschikking in de Zorg. Tilburg: SDU-uitgevers; 2018.
35. Irizarry T, DeVito DA, Curran CR. Patient portals and patient engagement: a state of the science review. J Med Internet Res 2015;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
36. Pisciotta M, Denneson LM, Williams HB, Woods S, Tuepker A, Dobscha SK. Providing mental health care in the context of online mental health notes: advice from patients and mental health clinicians. J Ment Health 2019 Feb;28(1):64-70. [doi: 10.1080/09638237.2018.1521924] [Medline: 30468100]
37. Dobscha SK, Kenyon EA, Pisciotta MK, Niederhausen M, Woods S, Denneson LM. Impacts of a web-based course on mental health clinicians' attitudes and communication behaviors related to use of OpenNotes. Psychiatr Serv 2019 Jun 01;70(6):474-479. [doi: 10.1176/appi.ps.201800416] [Medline: 30890047]
38. Rivard P. Doing Action Research in Your Own Organization. 5th edition. London: Sage Publications; 2002:533-534.

Abbreviations

EHR: electronic health record
NPT: normalization process theory

Edited by R Kukafka; submitted 18.02.21; peer-reviewed by M Hägglund, S Dobscha, L Petersson, G Myreteg; comments to author 23.04.21; revised version received 16.06.21; accepted 05.07.21; published 27.08.21.

Please cite as:
van Rijt AM, Hulter P, Weggelaar-Jansen AM, Ahaus K, Pluut B
Mental Health Care Professionals' Appraisal of Patients' Use of Web-Based Access to Their Electronic Health Record: Qualitative Study
J Med Internet Res 2021;23(8):e28045
URL: https://www.jmir.org/2021/8/e28045
doi:10.2196/28045
PMID:34448705

©Antonius Mattheus van Rijt, Pauline Hulter, Anne Marie Weggelaar-Jansen, Kees Ahaus, Bettine Pluut. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

The Application of a Case-Based Social Media–Assisted Teaching Method in Cariology Education Comparative Study

Li Li[1,2,3], MD; Xiaobin Liu[1,3], MD; Zeyuan Chen[1,3], BSc; Liyuan Wang[1,3], MD; Xiaoli Lian[1,3], MD; Huiru Zou[1,2,3], DDS

[1]Tianjin Key Laboratory of Oral and Maxillofacial Function Reconstruction, Tianjin Stomatological Hospital, The Affiliated Stomatological Hospital of Nankai University, Tianjin, China

[2]No 2 Teaching & Research Department of Conservative Dentistry, Endodontics and Oral Medicine, Tianjin Medical College, Tianjin, China

[3]School of Medicine, Nankai University, Tianjin, China

Corresponding Author:
Huiru Zou, DDS
Tianjin Key Laboratory of Oral and Maxillofacial Function Reconstruction
Tianjin Stomatological Hospital
The Affiliated Stomatological Hospital of Nankai University
No 75 Dagu Road
Heping District
Tianjin, 300041
China
Phone: 86 18222876129
Email: zouhuiru@163.com

Abstract

Background: Current cariology education based on the traditional teaching method faces a lot of challenges. Meanwhile, the COVID-19 pandemic caused an unprecedented disruption in medical education and health care systems worldwide. Innovation in the teaching mode of cariology education is required to change the situation.

Objective: The goal of the research was to evaluate the application effects of a case-based social media–assisted teaching method in cariology education.

Methods: Dental students of class 2019 were enrolled into the experimental group, while students of class 2018 served as control. A case-based social media–assisted teaching method was used in the experimental group, which included preclass activity via social media, additional discussion and practice process record in class, and questions and answers on the platform after class. The traditional teaching method, which consisted of conventional preparation before class, traditional lectures and demonstrations followed by students practice in class, and questions and answers step after class, was used in the control group. The teaching materials were the same in both groups. At the end of the program, students from both groups took cavity preparation skill evaluation tests. Questionnaires were tested on the case-based social media–assisted teaching group students anonymously. All data were analyzed using SPSS statistical software (version 22.0, IBM Corp).

Results: The mean student cavity preparation skill evaluation scores was 82.51 (SD 6.82) in the experimental group and 77.19 (SD 5.98) in the control group ($P<.05$). The questionnaire response rate was 100%. Of those, 94.3% (100/106) of the students recommended the case-based social media–assisted teaching method in cariology education. The majority of the participants agreed that it helped them memorize the theoretical knowledge of cariology, facilitated in-depth discussion, improved their enthusiasm and initiative in learning, and enhanced the relationship between teachers and students (104/106, 98.1%). They also recognized that the classroom atmosphere was active (94/106, 88.7%).

Conclusions: The case-based social media–assisted teaching method was beneficial in terms of learning, as demonstrated by the statistically significant improvement of the cavity preparation skill evaluation scores and satisfaction from attending students. This method could be used to supplement the teaching of cariology.

(J Med Internet Res 2021;23(8):e29372) doi:10.2196/29372

KEYWORDS

social media; case-based learning; cariology; dental cavity preparation; college students

Introduction

As one of the key subjects of clinical dentistry, cariology education plays an important role in dental education [1]. Students need to master not only theoretical knowledge but also operational skills in order to be able to resolve clinical problems encountered independently. Cariology education is still mainly based on the traditional teaching method, which faces challenges such as the demand of implementing didactic instruction into clinical training and the high occurrence of depression, anxiety, and stress encountered by students; the relationship between students and teachers need to be more interactive [2-5]. Meanwhile, the COVID-19 pandemic has caused an unprecedented disruption in medical education and health care systems worldwide [6]. Cariology education has also been greatly affected. An innovative teaching mode is required to improve teaching quality and cope with the impact of COVID-19 on teaching activities.

Case-based learning (CBL) is an effective teaching tool used in a variety of medical fields that links theory to practice by illustrating teaching points with actual clinical cases, promoting active and self-directed learning, clinical reasoning, and problem solving [7,8]. CBL has been encouraged by a number of universities and colleges. Most instructors are aware of the benefits of CBL, and as a result, the contents of courses are being reviewed and improved. Students appreciate that what is expected of them has been made clearer [9,10]. CBL has been used in Tianjin Medical College in cariology education for many years.

Nowadays, students use YouTube, Twitter, Instagram, and Facebook to access learning sources. Social media–assisted teaching in medical education appears to have increased and evolved during COVID-19 [11-14]. Information provided by these platforms is of varying quality, and it is hard for students to identify which sources are reliable. They need guidance from teachers, tutors, and supervisors. WeChat is a free app launched by Tencent in 2011 that has become one of the most popular communication tools in China, with more than 1 billion active users and coverage in more than 200 countries and regions worldwide. It provides instant messaging services; supports toll-free calls; allows sharing of audio files, pictures, and videos; and delivers other communication services. As new media, WeChat has an important impact on the daily life, learning, and interpersonal communications of students [15]. Previously, research on social media–assisted teaching focused on Twitter, Facebook, Second Life, and other platforms. Only a few studies recognized the educational function of WeChat [16,17]; teachers more often criticized students for spending time on smartphones and getting distracted. Using social media productively and achieving satisfactory results remains a big challenge [18].

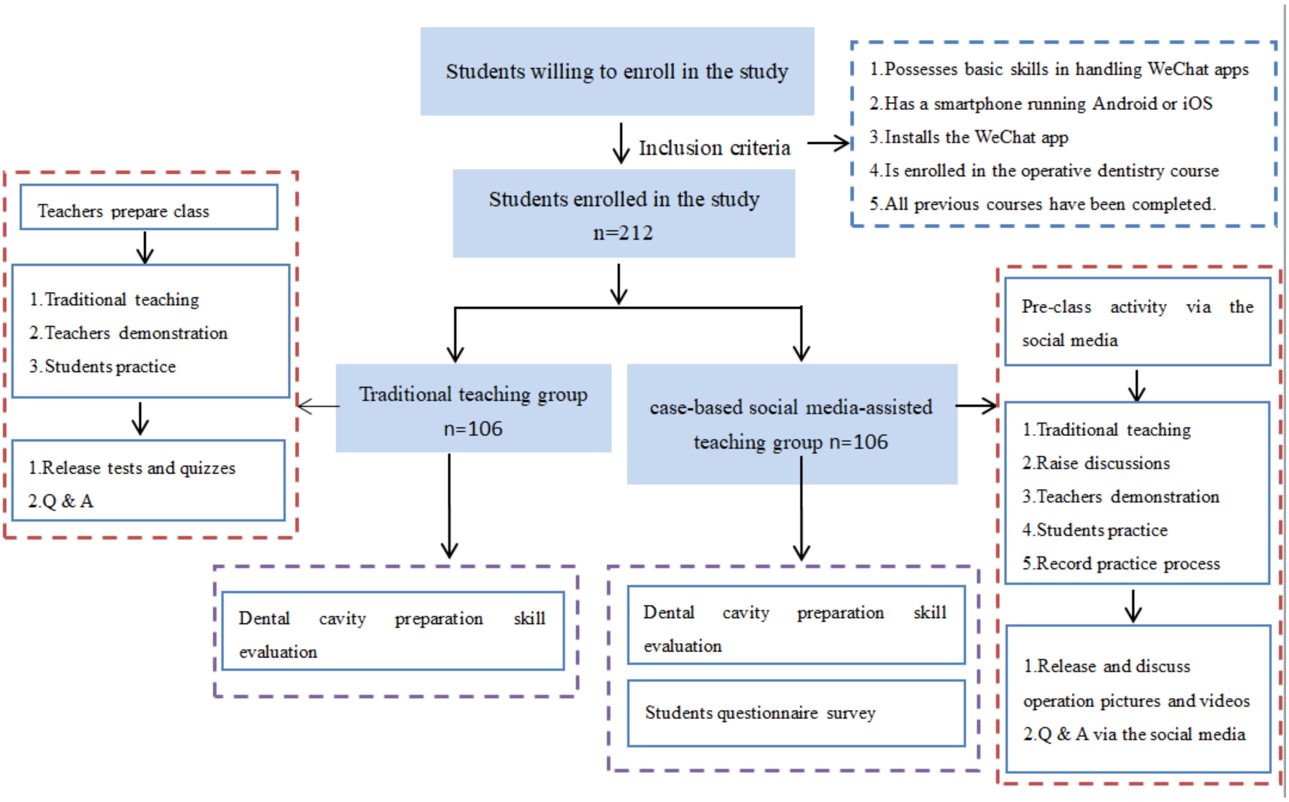
Therefore, this study aims to explore a case-based social media–assisted teaching method based on real clinical caries cases via WeChat for dental undergraduate cariology education and demonstrate its feasibility and acceptability.

Methods

Participant Criteria and Ethical Approval

Dental students in the 2018 and 2019 classes majoring in stomatology joined in this study. Inclusion criteria were (1) possess basic skills in using WeChat apps, (2) have a smartphone running either Android or iOS, (3) have the WeChat app installed, (4) be enrolled in the cariology course, and (5) have completed all previous courses. Exclusion criteria were on long-term leave or having dropped out for any reason. Figure 1 shows a flowchart of the process. For standardization, the teachers were all from the No. 2 Teaching and Research Department of Conservative Dentistry, Endodontics, and Oral Medicine, of the Affiliated Stomatological Hospital of Nankai University with standard training and calibration. Ethical approval for this study was obtained from the Medical Ethics Committee of Tianjin Medical College (ID E20180008). Informed consent was obtained from all students before the program started.

Figure 1. Flowchart of the process.


Students willing to enroll in the study
Inclusion criteria
1.Possesses basic skills in handling WeChat apps
2.Has a smartphone running Android or iOS
3.Installs the WeChat app
4.Is enrolled in the operative dentistry course
5.All previous courses have been completed.
Students enrolled in the study
n=212
Teachers prepare class
1.Traditional teaching
2.Teachers demonstration
3.Students practice
1.Release tests and quizzes
2.Q & A
Traditional teaching group
n=106
case-based social media-assisted
teaching group n=106
Pre-class activity via the social media
1.Traditional teaching
2.Raise discussions
3.Teachers demonstration
4.Students practice
5.Record practice process
1.Release and discuss operation pictures and videos
2.Q & A via the social media
Dental cavity preparation skill evaluation
Dental cavity preparation skill evaluation
Students questionnaire survey

Experimental Design

Students of class 2018 were taught using the traditional teaching method and served as the control group, while students of class 2019 were taught using the case-based social media–assisted teaching method and served as the experimental group. Data were collected during the first semester of the 2018-2019 and 2019-2020 academic years. Students were divided into several small WeChat groups. More than 20 cases and problems on different topics were prepared by the faculty in the department. The teacher presented a brief explanation, followed by an introduction of a clinical case scenario accompanied by questions addressing the objectives of that part of the lecture. The students were asked to analysis the cases and make treatment plans accordingly. The experimental group used social media (mainly WeChat) within the whole training program. First, the teachers set up the experimental group before class using smartphones and invited students to join the group by releasing the group QR code. In addition to the same steps as the conventional preclass preparation of the traditional teaching group, the teachers would post plenty of pictures and short videos combined with text descriptions on the platform. Meanwhile, the teaching PowerPoint slides and preclass thinking questions would be released to the students. Second, the teachers reviewed the theoretical knowledge with students in class and led discussions about the preclass pictures and videos. The students then practiced according to the teachers' instructions. An assistant teacher recorded and took pictures of the students' practical positions, typical errors, and the cavity models prepared by the students. Finally, the teacher made comments and summarized after students' practice. Third, the pictures and videos recorded by the assistant teacher would be released in the experimental group after class to inspire the students to discuss and work out problems, and the teachers would make comments. Excellent practice pictures and videos would be highlighted and shared with students. Students could learn independently, and peer evaluation was encouraged in order to stimulate students' enthusiasm and participation. At the same time, the questions raised by students at any time would be answered through the platform by teachers.

The control group students experienced teaching activities according to the traditional teaching model. The same cases and problems of different topics were prepared by the faculty in the department and presented to students. Details were as follows: First, the teachers prepared according to the textbooks and syllabus before class and made teaching PowerPoint slides. Second, the teachers reviewed the theoretical knowledge in class first followed by video and live demonstrations. The students then practiced according to the teachers' instructions, and the teachers made comments and summarized after the students' practice. Third, teachers released tests and quizzes after class that required students to finish within certain times and answered the questions raised by students.

Cavity Preparation Skill Evaluation

All students were given the same introductory lecture via PowerPoint presentation and the same demonstration from one trained teacher on the design and instrumentation of conventional class I cavity preparation. After that, students were asked to prepare class I cavities on the standard laser scanning model lower left first molar teeth (UNISN Inc). One tooth was prepared by each student within the 40 minute class. The teachers would then evaluate their cavity preparation skill levels using the Cavity Preparation Skill Evaluation System (CPSES). Briefly, the CPSES system scanned, evaluated, and scored the outline form and depth of prepared cavities against a theoretically ideal tooth model mounted on a special jig. The laser beam image was taken with 2 specialized digital cameras and processed with an image processing system. The converted image was transferred onto computer, and the contour and depth of the cavity was calculated. Finally both the 3D images of the prepared tooth and ideal tooth were displayed on the screen with the score and evaluation details, which could be printed out.

Student Evaluation Survey

Meanwhile, in order to qualitatively analyze the impact of the case-based social media–assisted method on cariology education, student evaluation surveys were used as one of the main assessments. Interviews and group discussions were also employed in this study in order to get firsthand feedback and true voices from students. Examples of the questions were as follows: Are you satisfied with the case-based social media–assisted teaching method in cariology education? Are you interested in learning cariology? Did the case-based social media–assisted teaching method help you memorize caries-related knowledge and integrate the theoretical knowledge with clinical practice? The participants were asked to answer the questions anonymously at the end of the term.

Statistical Analysis

All data were collected and analyzed with SPSS statistical software (version 22.0, IBM Corp). The Levene test was used to assess variance homogeneity for the paired *t* test. The Kolmogorov-Smirnov test was used to evaluate the normality of the sample distribution. The level of statistical significance was set at α=.05. The results of practical training were expressed as mean and standard deviation, and the data of the 2 groups were analyzed by paired *t* test. *P*<.05 indicated that the difference was statistically significant.

Results

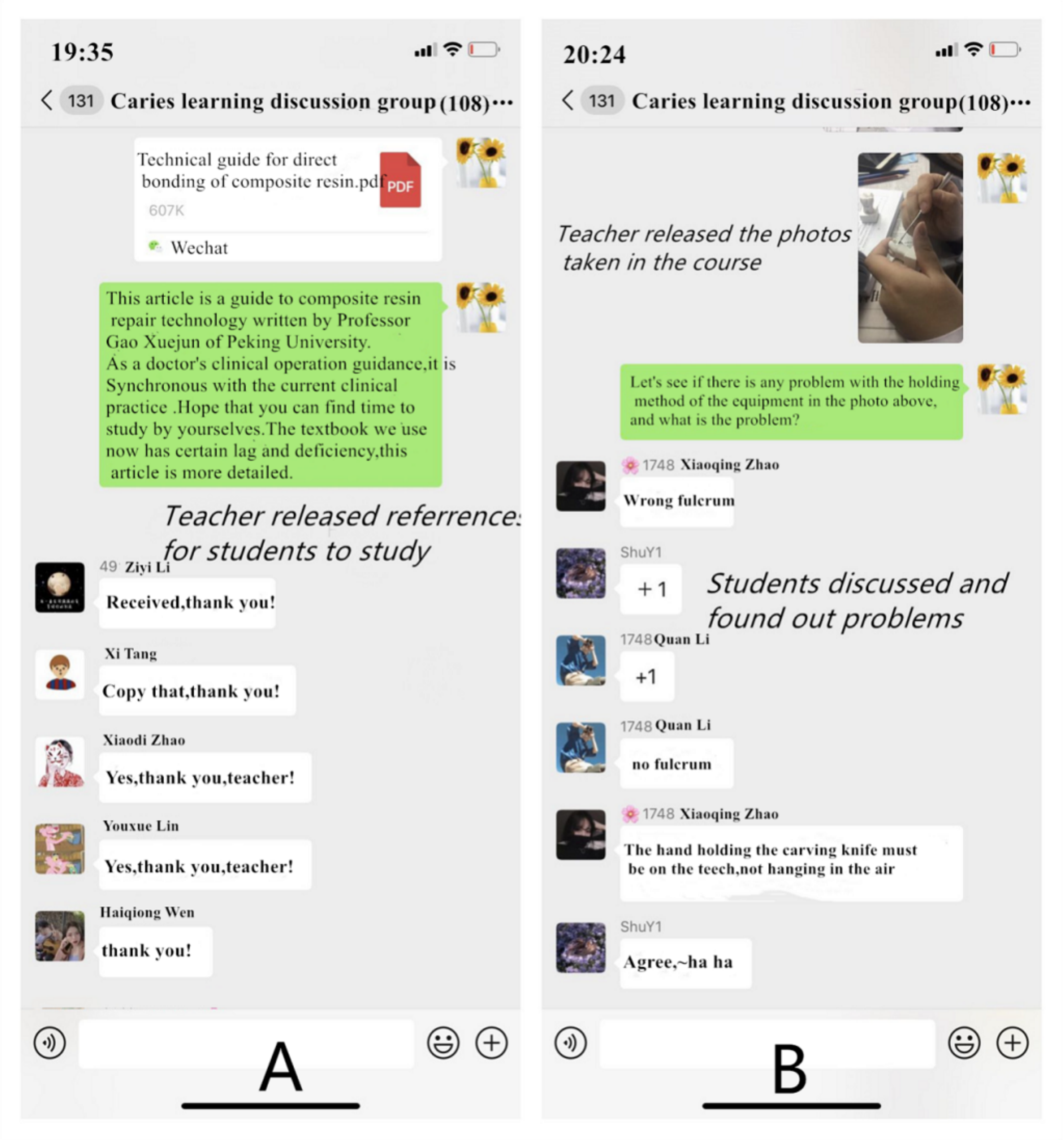
Participant Analysis

There were 106 students each in class 2018 taught using the traditional teaching method, serving as the control group and class 2019 taught with the case-based social media–assisted teaching method, serving as the experimental group. The mean age of the students was 19.76 (SD 2.74) years in the experimental group versus 20.00 (SD 3.98) years in the control group (*P*=.97). A total of 78.3% (83/106) in the experimental group were women compared to 70.8% (75/106) in the control group (*P*=.39). No significant difference was seen between the 2 groups in terms of students' previous scores (*P*=.69).

Teacher-Student Interaction Via WeChat

Figure 2 shows the teacher-student interactions based on the WeChat social media platform. Pictures and short videos combined with text descriptions of clinical cases were presented and discussed on the platform. Over the period of the first semester of the 2019-2020 academic year, hundreds of chat-chat discussions took place on the WeChat social media platform. Notifications such as course notices, preclass questions, after-class study materials, and questionnaires were posted on this platform. Students felt that whenever they posted any question, there would be answers from the teachers or their classmates immediately. The WeChat learning tool was portable and instantly accessible.

Figure 2. Teacher-student interaction on the social media platform: (A) teacher-released references for students' self study and (B) teacher-released pictures and videos taken in the course with student discussion and resolution of problems.


19:35
131 Caries learning discussion group(108)···
PDF
607K
Wechat
A
20:24
131 Caries learning discussion group(108)···
B

Technical guide for direct bonding of composite resin.pdf

This article is a guide to composite resin repair technology written by Professor Gao Xuejun of Peking University. As a doctor's clinical operation guidance,it is Synchronous with the current clinical practice .Hope that you can find time to study by yourselves.The textbook we use now has certain lag and deficiency,this article is more detailed.

Teacher released referrences for students to study

49 Ziyi Li: Received,thank you!

Xi Tang: Copy that,thank you!

Xiaodi Zhao: Yes,thank you,teacher!

Youxue Lin: Yes,thank you,teacher!

Haiqiong Wen: thank you!

Teacher released the photos taken in the course

Let's see if there is any problem with the holding method of the equipment in the photo above, and what is the problem?

1748 Xiaoqing Zhao: Wrong fulcrum

ShuY1: + 1

Students discussed and found out problems

1748 Quan Li: +1

1748 Quan Li: no fulcrum

1748 Xiaoqing Zhao: The hand holding the carving knife must be on the teech,not hanging in the air

ShuY1: Agree,~ha ha

Cavity Preparation Skill Evaluation Results

At the end of the training course, the dental cavities prepared by the students of both groups were assessed by CPSES (Figure 3). The results show that the scores of the experimental group were higher than those of the traditional teaching group. The scores of the experimental group were 82.51 (SD 6.82) and those of the control group were 77.19 (SD 5.98). There was a statistically significant difference between these groups ($P<.05$). Figure 4 shows the assessment results of the outline form and depth revealed by CPSES. There were significant improvements in the cavity outline form, retention form, smoothness, cavity depth, and cavity margin angulations of the experimental group-prepared dental cavities compared with the control group.

Figure 3. Cavity preparation skill evaluation system (CPSES).

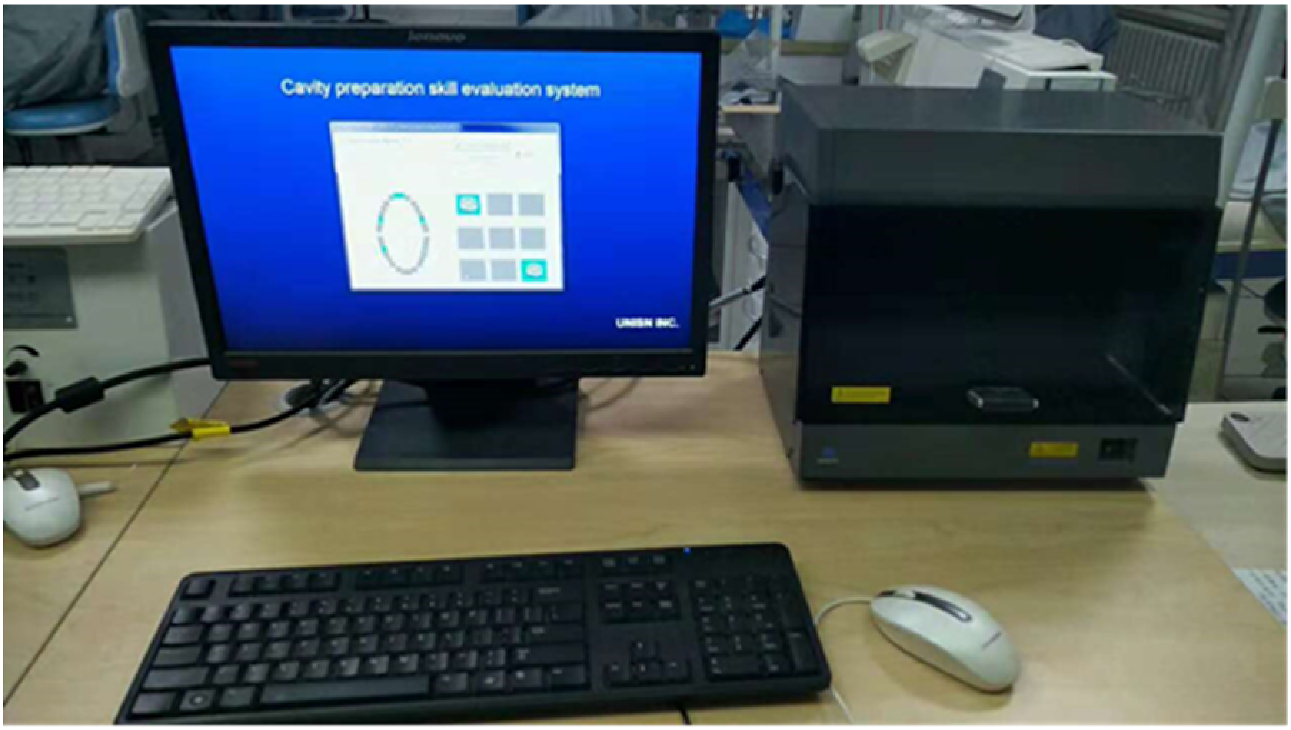
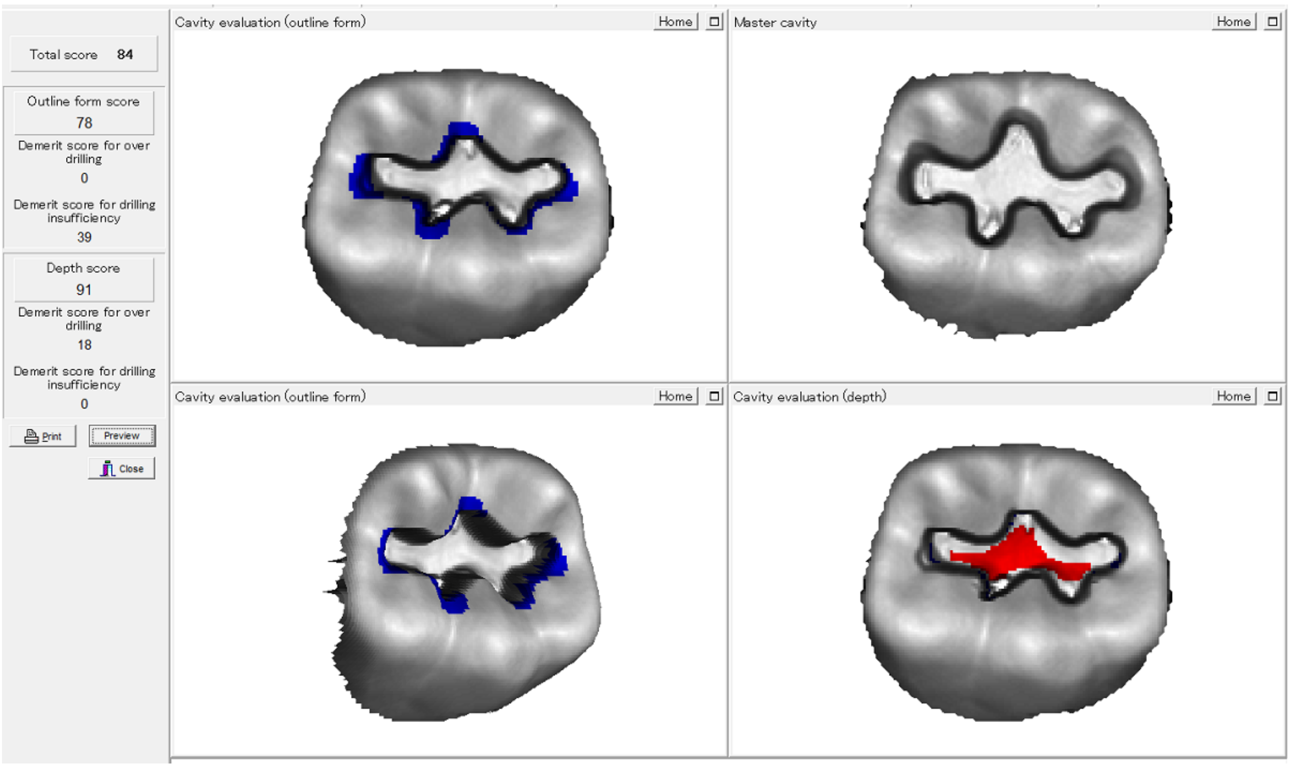

Cavity preparation skill evaluation system
UNISN INC.

Figure 4. Assessment result revealed by cavity preparation skill evaluation system.



Total score 84

Outline form score
78

Demerit score for over drilling
0

Demerit score for drilling insufficiency
39

Depth score
91

Demerit score for over drilling
18

Demerit score for drilling insufficiency
0

Print
Preview
Close
Cavity evaluation (outline form)
Home
Master cavity
Home
Cavity evaluation (outline form)
Home
Cavity evaluation (depth)
Home

Student Evaluation Survey Results

Completed questionnaires were returned from all 106 students of the experimental group. Thus, the response rate was 100%. A total of 94.3% (100/106) of the students were interested in the teaching method. The case-based social media–assisted teaching method improved students' interest in learning, which helped them clearly understand each knowledge point, understand and memorize related knowledge, and master the key points of cavity preparation skills. It improved students' enthusiasm and initiative in learning, and enhanced the relationship between teachers and students, as shown in Table 1. The majority of the interview and group discussion results were positive about the use of the case-based social media–assisted teaching method in cariology courses, which also confirmed the questionnaire results.

Table 1. Evaluation of teaching effect of experimental group.

Question	Yes		No	
	n	Constituent ratio (%)	n	Constituent ratio (%)
Are you satisfied with the case-based social media–assisted teaching method in cariology and restorative dentistry education?	100	94.3	6	5.7
Did the WeChat-assisted teaching method help you memorize related knowledge and integrate the theoretical knowledge with clinical practice?	104	98.1	2	1.9
Are you interested in learning cariology and restorative dentistry?	90	84.9	16	15.1
Was the classroom atmosphere active?	94	88.7	12	11.3
Are you satisfied with the degree of mastering cariology and restorative dentistry?	98	92.5	8	7.5
Did the teacher explain the material clearly?	96	90.6	10	9.4
Did the case-based social media–assisted teaching method help you understand the various knowledge points and memorize related knowledge?	106	100	0	0
Did the case-based social media–assisted teaching method improve learning enthusiasm and initiative?	104	98.1	2	1.9
Did the case-based social media–assisted teaching method enhance the relationship between teachers and students?	104	98.1	2	1.9

Discussion

Cariology Education Requirements

Dental caries are one of the most prevalent chronic diseases in the world. They have historically been considered the most important global oral health burden [19]. They are still a major health problem in many countries. Because of their high incidence rate, low treatment rate, high retreatment rate, and close relationship with general health, special attention still needs to be given in terms of prevention and treatment. Thus, the demand for delivering a quality cariology education for dental students is increasing.

Even though cariology is being taught in different ways in different countries, it is generally agreed that students should on graduation have a sound theoretical knowledge and understanding of cariology together with an adequate clinical experience to be able to resolve clinical problems encountered independently and without assistance. A professional attitude and behavior is also required. Students need to learn how to establish a trusting patient-dentist relationship and how to work with other members of the dental team [20,21]. Therefore students should continuously develop those skills including a thorough understanding and application of their moral, ethical, and legal responsibilities. CBL provides students patient details such as clinical signs and symptoms, a history of present illness, past history, and laboratory investigations. Students are actively involved in the discussion, interact with each other, and work together through the social media platform. CBL allows teachers more input into the direction of learning and induces learning on a deeper level. It allows students to be exposed to real-life case scenarios and enhances their analytical and thinking processes by this student-centered learning approach [22]. Cariology education is changing and adapting quickly to new technologies, and this has an impact directly on practical results.

CBL Enhanced Theoretical-Clinical Integration and Communication

In this study, a case-based social media–assisted teaching method was applied to the cariology course. The effects were compared with a previous traditional teaching class. Among the students from the experimental group, 98.1% recognized that the case-based social media–assisted teaching method helped them memorize caries-related knowledge and integrate the theoretical knowledge with clinical practice. CBL improved the ability of the students to solve clinical problems, focusing on the integration of basic knowledge to clinical practice in the context of case-based learning [23]. At the same time, communication on the social platform between teachers and students encouraged students and increased their participation opportunities, providing personalized feedback according to the students' performances. Previous research had stressed that knowing how much a student was involved in creating a learning environment would help teachers understand the student's preference [24]. In another study, students were divided into several WeChat groups with 3 to 4 students per group. The teacher collected clinical data of patients in electronic records, images, and videos and delivered the materials to the WeChat groups before class with several questions. The students were asked to study the material individually and discuss the questions with each other in their small groups [25]. In this way, students could communicate and exchange ideas with their peers to solve the problems. Teachers could also share their ideas and experiences on the platform. WeChat became a communication tool for both teachers and students and also a platform for group work. Through this, core skills of evidence-based dental practice within the curriculum were highlighted. Students were able to identify uncertainty or gaps in understanding and formulating answerable questions, search for evidence using appropriate resources, and search for and use the most appropriate clinical guidelines. Teachers were able to deliver lifelong learning skills to the students.

CBL Promoted Dental Cavity Preparation Skills

Dental cavity preparation and restoration technology involved in cariology are fundamental and closely related to clinical work. The traditional teaching mode of a practical training course mostly adopts the teaching mode of "teacher's demonstration, student's practice, teacher's comments." Due to the narrow space in the oral cavity, it is difficult for every student to observe the teacher's demonstrations on the phantom head clearly. Additionally, students cannot effectively recognize the problems in their own operation. At the same time, due to the limitation of teaching time, teachers' comments are often short, general, and cannot cover all aspects. The end of the training course means the end of the learning of practical skills. It is difficult for students to understand the biological, medical, basic, and applied clinical sciences in such a short period and apply knowledge to recognize caries and other dental hard tissue disorders and make decisions about their prevention and management. The social media–assisted teaching method is very helpful in carrying out teaching demonstrations. Before class, the teacher records the key operation steps and precautions on small videos or takes pictures for students to preview. This resolves time and space limitations of traditional teaching and learning methods, cultivates students' independent learning ability, and meets the personalized learning needs of different levels of students. The theoretical knowledge test scores of social media–assisted teaching groups were significantly higher than those of traditional teaching groups. Previous studies also showed that social media had a generally positive impact on teaching effect, consistent with the results of this study [17,18,26]. However, most of those studies only used theoretical tests, questionnaires, students' feedback to evaluate the teaching effects, lacking objective evaluation indicators. In recent years, we have adopted CPSES to make assessments through the computer 3D scanning device and image processing system. The results are objective and fair, excluding the interference of subjective factors, making the assessments more scientific [27,28]. This study continued to use this evaluation system to objectively evaluate the effect of social media–assisted teaching in the dental cavity preparation skill training. The results showed that students' cavity preparation scores in the social media–assisted teaching group were significantly higher than those in the traditional teaching group.

CBL Extended Traditional Teaching

Social media–assisted teaching was a powerful supplement to the traditional classroom teaching, which significantly enhanced the teaching effect. After class, the typical classroom operation videos and pictures taken by the assistant teacher were sent to the students immediately. Students had a clearer understanding of their own problems and learned through the strengths and weaknesses of others. Yang et al [29] conducted a study on teaching fundamental nursing skills with smartphone-based video feedback to facilitate teaching. Results showed that video feedback was an effective way to improve nursing students' academic performance and professional skills. In our study, in the process of gypsum cavity preparation and phantom head cavity preparation, the operation fulcrum was emphasized. Students were required to have a stable fulcrum when they operated. Beginners often easily ignored this problem. In the social media–assisted teaching group, teachers sent pictures of students practicing without a fulcrum or with a wrong fulcrum to the students after class. This action formed a strong visual stimulation for students and enhanced their memory about this point. Teachers asked questions and encouraged students to discuss and collaborate with others. Some studies have confirmed that online teaching methods could enable students to get timely feedback from teachers and other students in the learning process, and students were familiar with online communication methods and participated in discussions more actively [13,29]. Teaching social and communicative competencies has become an important part of undergraduate dental education [20]. Studies have also shown that social media promoted the formation of a student-centered teaching mode by providing information and interactive feedback anytime and anywhere, which significantly improves the proportion of students actively participating in learning and enhanced the satisfaction of students [13]. These results were also confirmed in our study. As the students had rarely been exposed to an activity like CBL, it was observed that some students participated actively while some remained passive in the traditional method [30]. In this study, the WeChat platform made the learning mode more diversified, enabled students to grasp the key points of practical training more solidly, and effectively extended the traditional teaching. The platform also reduced the anxiety and tension of face-to-face communication between students and teachers, and made communication occur more smoothly. WeChat was helpful in eliminating the possible barriers between teachers and students and effectively enhanced the relationships between teachers and students. The results of our study showed that students were actively participating in the whole teaching and learning process, communicating and interacting with peers as well as faculty members on the social platform. Meanwhile, extra financial costs of maintaining and updating modules were not required, which made it more popular and likely to be applied in other fields.

Case-based social media–assisted teaching in cariology education promoted students' learning and mastering of advanced treatment technology and the expansion of cutting-edge knowledge in clinical treatment. Teachers used social media to provide multimedia teaching materials, such as pictures, audios, and videos, so that students could learn in an all-round way more intuitively and vividly through the network anytime and anywhere, providing a convenient way for teaching [31]. In order to better adapt to the needs of clinical technology development, cariology teaching should also constantly update the knowledge system so that students can use the latest medical equipment, materials, and methods. Teachers can use social media to demonstrate advanced treatment technology to students (eg, minimally invasive comprehensive prevention and treatment of caries, preparation and filling technology of cavities under the microscope) so that students' learning is not out of line with clinical practice, which makes up for the slight lag of current textbooks. Most of the students in medical colleges have not received systematic scientific research training. They lack the ability to access knowledge effectively through the internet. Teachers could push new authorized theoretical knowledge directly to students on social media so that students can learn new knowledge more effectively. Through the study of current advanced knowledge and technology, students can realize the difference between practical training operation and current clinical operation. They can also discuss the difference between practical training operation technology and clinical advanced technology, which will stimulate students' learning initiative and enthusiasm, improving the learning effects. The improvement of the learning effect would certainly improve the competitiveness of students. Several studies have confirmed that the use of social media promoted development of effective work relationships and were helpful for a successful professional career [32,33]. Inspired by the clinical cases delivered by teachers through social media, students gained knowledge of the impact of restorative procedures on mucosa, periodontal tissues, occlusion, and oral function. They also learned the skills of selecting and handling appropriate restorative materials considering physical and chemical properties, biocompatibility, and longevity and selecting and carrying out operative techniques appropriate for both material and case. Through clinical cases introduced on social media, students achieved skills of deciding when, how, and to what extent to remove carious tissue before the placement of a restoration, considering the restorability of the tooth, preservation of tooth structure, and pulp vitality. Meanwhile, they were also aware of the health economic aspects of restorative therapy and patient privacy protection issues. It was found that social media–assisted teaching improved students' independent learning, learning interest, enthusiasm, and ability to cooperate with peers, thus promoting education reform [34]. Therefore, students would have a perceptual understanding of the training courses to be learned.

Limitations

First, the study population was restricted to one medical college in China. Therefore, it might limit the generalizability of the study. Further studies carried out in other colleges or universities would be expected, which would provide stronger evidence on the effectiveness of the case-based social media–assisted teaching method.

Another limitations of the study was the lack of randomized controlled design. The premises were that education in experimental design should be undertaken with an awareness of educational principles. Unlike when using laboratory animals, more restricted ethical considerations needed to be considered. Students deserved to learn with the most suitable and most effective teaching method. Therefore, we decided to enroll the entire 2019 class of students into the experimental group and chose the previous class with the traditional teaching method as control. Since students knew very little about cariology and had no capability to prepare dental cavities before class, the pretest posttest quasi-experiments were not carried out in this study. Hence, the results of this study should be interpreted with caution. We hope that the case-based social media–assisted teaching method will be used to help students be equipped to provide appropriate care for the most widespread oral disease.

Conclusion

The case-based social media–assisted teaching method was beneficial in cariology education in terms of learning, as demonstrated by the statistically significant improvement of CPSES and satisfaction of the attending students. The application of the case-based social media–assisted teaching method in cariology education helped students uptake the theoretical knowledge, ignited their enthusiasm and motivation in learning, and relieved their stress. The pilot study revealed that the case-based social media–assisted teaching method could be used to supplement the teaching of cariology and is worth further study and application.

Acknowledgments

This study was funded by grant 2018B-N02028 from the Medical Education Research Project of Medical Education Branch of Chinese Medical Association, grant FIB180534 from the National Office for Education Sciences 2018 Planning Project of 13th Five-Year Scientific Research, and grant ZD20016 of the Tianjin Health and Family Planning Commission of the People's Republic of China.

Conflicts of Interest

None declared.

References

1. Hilton TJ. Summitt's Fundamentals of Operative Dentistry: A Contemporary Approach. Chicago: Quintessence Publishing Co; 2013.
2. Tikhonova S, Girard F, Fontana M. Cariology education in Canadian dental schools: Where are we? Where do we need to go? J Dent Educ 2018 Jan;82(1):39-46. [doi: 10.21815/JDE.018.006] [Medline: 29292324]
3. Raphael SL, Foster Page LA, Hopcraft MS, Dennison PJ, Widmer RP, Evans RW. A survey of cariology teaching in Australia and New Zealand. BMC Med Educ 2018 Apr 10;18(1):75 [FREE Full text] [doi: 10.1186/s12909-018-1176-4] [Medline: 29631580]
4. Basudan S, Binanzan N, Alhassan A. Depression, anxiety and stress in dental students. Int J Med Educ 2017 May 24;8:179-186 [FREE Full text] [doi: 10.5116/ijme.5910.b961] [Medline: 28553831]
5. Oderinu OH, Adegbulugbe IC, Orenuga OO, Butali A. Comparison of students' perception of problem-based learning and traditional teaching method in a Nigerian dental school. Eur J Dent Educ 2020 May 06;24(2):207-212. [doi: 10.1111/eje.12486] [Medline: 31765048]

6. Tolsgaard MG, Cleland J, Wilkinson T, Ellaway RH. How we make choices and sacrifices in medical education during the COVID-19 pandemic. Medical Teacher 2020 May 22;42(7):741-743. [doi: 10.1080/0142159x.2020.1767769]
7. Kaur G, Rehncy J, Kahal KS, Singh J, Sharma V, Matreja PS, et al. Case-based learning as an effective tool in teaching pharmacology to undergraduate medical students in a large group setting. J Med Educ Curric Dev 2020 May 13;7:2382120520920640 [FREE Full text] [doi: 10.1177/2382120520920640] [Medline: 32435693]
8. Alhazmi A, Quadri MFA. Comparing case-based and lecture-based learning strategies for orthodontic case diagnosis: a randomized controlled trial. J Dent Educ 2020 Aug 06;84(8):857-863. [doi: 10.1002/jdd.12171] [Medline: 32374442]
9. Alsunni AA, Rafique N. Effectiveness of case-based teaching of cardiovascular physiology in clinical pharmacy students. J Taibah Univ Med Sci 2021 Feb;16(1):22-28 [FREE Full text] [doi: 10.1016/j.jtumed.2020.11.009] [Medline: 33603628]
10. Jacob SA, Dhing OH, Malone D. Perceptions of Australian and Malaysian educators in an undergraduate pharmacy program on case-based learning. Am J Pharm Educ 2019 Apr 01;83(3):6597 [FREE Full text] [doi: 10.5688/ajpe6597] [Medline: 31065163]
11. Wanner GK, Phillips AW, Papanagnou D. Assessing the use of social media in physician assistant education. Int J Med Educ 2019 Jan 29;10:23-28 [FREE Full text] [doi: 10.5116/ijme.5c14.ef82] [Medline: 30694797]
12. Maggio LA, Leroux TC, Artino AR. To tweet or not to tweet, that is the question: a randomized trial of Twitter effects in medical education. PLoS One 2019 Oct 16;14(10):e0223992 [FREE Full text] [doi: 10.1371/journal.pone.0223992] [Medline: 31618267]
13. Javaeed A, Kibria Z, Khan Z, Ghauri SK. Impact of social media integration in teaching methods on exam outcomes. Adv Med Educ Pract 2020 Jan;11:53-61. [doi: 10.2147/amep.s209123]
14. Laurentino Lima D, Nogueira Cordeiro Laurentino Lima R, Benevenuto D, Soares Raymundo T, Shadduck PP, Melo Bianchi J, et al. Survey of social media use for surgical education during COVID-19. JSLS 2020;24(4):e2020.00072. [doi: 10.4293/jsls.2020.00072]
15. Wang J, Gao F, Li J, Zhang J, Li S, Xu G, et al. The usability of WeChat as a mobile and interactive medium in student-centered medical teaching. Biochem Mol Biol Educ 2017 Sep 05;45(5):421-425 [FREE Full text] [doi: 10.1002/bmb.21065] [Medline: 28581156]
16. Luan H, Wang M, Sokol RL, Wu S, Victor BG, Perron BE. A scoping review of WeChat to facilitate professional healthcare education in Mainland China. Med Educ Online 2020 Dec 23;25(1):1782594 [FREE Full text] [doi: 10.1080/10872981.2020.1782594] [Medline: 32573367]
17. Zhang W, Li Z, Li Z. WeChat as a platform for problem-based learning in a dental practical clerkship: feasibility study. J Med Internet Res 2019 Mar 19;21(3):e12127 [FREE Full text] [doi: 10.2196/12127] [Medline: 30888328]
18. Li B, Wu Y, Jiang S, Zhai H. WeChat addiction suppresses the impact of stressful life events on life satisfaction. Cyberpsychol Behav Soc Netw 2018 Mar;21(3):194-198. [doi: 10.1089/cyber.2017.0544] [Medline: 29638153]
19. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018;392(10159):1789-1858. [doi: 10.3410/f.731220250.793569875]
20. Meyer-Lueckel H, Opdam NJM, Breschi L, Buchalla W, Ceballos L, Doméjean S, et al. EFCD Curriculum for undergraduate students in Integrated Conservative Oral Healthcare (ConsCare). Clin Oral Investig 2019 Sep 3;23(9):3661-3670. [doi: 10.1007/s00784-019-02978-x] [Medline: 31270666]
21. Lichtenstein NV, Haak R, Ensmann I, Hallal H, Huttenlau J, Krämer K, et al. Does teaching social and communicative competences influence dental students' attitudes towards learning communication skills? A comparison between two dental schools in Germany. GMS J Med Educ 2018;35(2):Doc18 [FREE Full text] [doi: 10.3205/zma001165] [Medline: 29963608]
22. Kaur G, Rehncy J, Kahal KS, Singh J, Sharma V, Matreja PS, et al. Case-based learning as an effective tool in teaching pharmacology to undergraduate medical students in a large group setting. J Med Educ Curric Dev 2020 May 13;7:2382120520920640 [FREE Full text] [doi: 10.1177/2382120520920640] [Medline: 32435693]
23. Bi M, Zhao Z, Yang J, Wang Y. Comparison of case-based learning and traditional method in teaching postgraduate students of medical oncology. Medical Teacher 2019 Jun 19;41(10):1124-1128. [doi: 10.1080/0142159x.2019.1617414]
24. Pyle E, Hung W. The role of subject presence type on student motivation in a PBL learning environment. Adv Health Sci Educ Theory Pract 2019 Oct 23;24(4):643-663. [doi: 10.1007/s10459-019-09889-2] [Medline: 31016430]
25. Zeng F, Deng G, Wang Z, Liu L. WeChat: a new clinical teaching tool for problem-based learning. Int J Med Educ 2016 Apr 25;7:119-121 [FREE Full text] [doi: 10.5116/ijme.5708.e5c4] [Medline: 27111920]
26. Tu S, Yan X, Jie K, Ying M, Huang C. WeChat: an applicable and flexible social app software for mobile teaching. Biochem Mol Biol Educ 2018 Sep 26;46(5):555-560 [FREE Full text] [doi: 10.1002/bmb.21170] [Medline: 30369037]
27. Zou H, Jin S, Sun J, Dai Y. A cavity preparation evaluation system in the skill assessment of dental students. J Dent Educ 2016 Aug 01;80(8):930-937. [doi: 10.1002/j.0022-0337.2016.80.8.tb06173.x]
28. Zou H, Jin S, Sun J, Dai Y. Cavity preparation skill evaluation system for assessing junior dental students. Med Educ 2016 May 13;50(5):586-587. [doi: 10.1111/medu.13041] [Medline: 27072475]

29. Yang X, Xie R, Chen S, Yu W, Liao Y, Krewski D, et al. Using video feedback through smartphone instant messaging in fundamental nursing skills teaching: observational study. JMIR Mhealth Uhealth 2019 Sep 05;7(9):15386 [FREE Full text] [doi: 10.2196/15386] [Medline: 31489839]
30. Gupta K, Arora S, Kaushal S. Modified case based learning: our experience with a new module for pharmacology undergraduate teaching. Int J App Basic Med Res 2014;4(2):90. [doi: 10.4103/2229-516x.136786]
31. Montag C, Becker B, Gan C. The multipurpose application WeChat: a review on recent research. Front Psychol 2018 Dec 11;9:2247 [FREE Full text] [doi: 10.3389/fpsyg.2018.02247] [Medline: 30618894]
32. Vande VL, Ryder H, Best J. Maximizing career advancement during the COVID-19 pandemic: recommendations for postgraduate training programs. Acad Med 2021 Jan:19. [doi: 10.1097/acm.0000000000003938]
33. Rajeh MT, Sembawa SN, Nassar AA, Al Hebshi SA, Aboalshamat KT, Badri MK. Social media as a learning tool: dental students' perspectives. J Dent Educ 2021 Apr 20;85(4):513-520. [doi: 10.1002/jdd.12478] [Medline: 33219515]
34. Guraya SY, Al-Qahtani MF, Bilal B, Guraya SS, Almaramhy H. Comparing the extent and pattern of use of social networking sites by medical and non medical university students: a multi-center study. Psychol Res Behav Manag 2019 Jul;12:575-584. [doi: 10.2147/prbm.s204389]

Abbreviations

CBL: case-based learning
CPSES: Cavity Preparation Skill Evaluation System

Edited by R Kukafka; submitted 05.04.21; peer-reviewed by JS Tham, G Halasz; comments to author 25.05.21; revised version received 07.06.21; accepted 05.07.21; published 13.08.21.

Please cite as:
Li L, Liu X, Chen Z, Wang L, Lian X, Zou H
The Application of a Case-Based Social Media–Assisted Teaching Method in Cariology Education：Comparative Study
J Med Internet Res 2021;23(8):e29372
URL: https://www.jmir.org/2021/8/e29372
doi:10.2196/29372
PMID:34397390

©Li Li, Xiaobin Liu, Zeyuan Chen, Liyuan Wang, Xiaoli Lian, Huiru Zou. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

The Impact of Gamification-Induced Users' Feelings on the Continued Use of mHealth Apps: A Structural Equation Model With the Self-Determination Theory Approach

Tong Wang[1], BS; Lingye Fan[2], MD; Xu Zheng[3], MSc; Wei Wang[1], PhD; Jun Liang[4], MS; Kai An[3], MSc; Mei Ju[5], MS; Jianbo Lei[6,7,8], MD, PhD

[1]Department of Medical Informatics, School of Public Health, Jilin University, Changchun, China
[2]Department of Obstetrics and Gynecology, Affiliated Hospital, Southwest Medical University, Luzhou, China
[3]Peking University Third Hospital, Beijing, China
[4]IT Center, Second Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, China
[5]School of Nursing, Southwest Medical University, Luzhou, China
[6]Institute of Medical Technology, Health Science Center, Peking University, Beijing, China
[7]Center for Medical Informatics, Health Science Center, Peking University, Beijing, China
[8]School of Medical Informatics and Engineering, Southwest Medical University, Luzhou, China

Corresponding Author:
Jianbo Lei, MD, PhD
Institute of Medical Technology, Health Science Center
Peking University
38 Xueyuan Rd, Haidian District
Beijing
China
Phone: 86 82805901
Email: jblei@hsc.pku.edu.cn

Abstract

Background: Continued use of mHealth apps can achieve better effects in health management. Gamification is an important factor in promoting users' intention to continue using mHealth apps. Past research has rarely explored the factors underlying the continued use of mobile health (mHealth) apps and gamification's impact mechanism or path on continued use.

Objective: This study aimed to explore the factors influencing mHealth app users' intention to continue using mHealth apps and the impact mechanism and path of users' feelings induced by gamification on continued mHealth app use.

Methods: First, based on the expectation confirmation model of information system continuance, we built a theoretical model for continued use of mHealth apps based on users' feelings toward gamification. We used self-determination theory to analyze gamification's impact on user perceptions and set the resulting feelings (competence, autonomy, and relatedness) as constructs in the model. Second, we used the survey method to validate the research model, and we used partial least squares to analyze the data.

Results: A total of 2988 responses were collected from mHealth app users, and 307 responses were included in the structural equation model after passing the acceptance criteria. The intrinsic motivation for using mHealth apps is significantly affected by autonomy (β=.312; P<.001), competence (β=.346; P<.001), and relatedness (β=.165; P=.004) induced by gamification. The intrinsic motivation for using mHealth apps has a significant impact on satisfaction (β=.311, P<.001) and continuance intention (β=.142; P=.045); furthermore, satisfaction impacts continuance intention significantly (β=.415; P<.001). Confirmation has a significant impact on perceived usefulness (β=.859; P<.001) and satisfaction (β=.391; P<.001), and perceived usefulness has a significant impact on satisfaction (β=.269; P<.001) and continuance intention (β=.273; P=.001). The mediating effect analysis showed that in the impact path of the intrinsic motivation for using the mHealth apps on continuance intention, satisfaction plays a partial mediating role (β=.129; P<.001), with a variance accounted for of 0.466.

Conclusions: This study explored the impact path of users' feelings induced by gamification on the intention of continued mHealth app use. We confirmed that perceived usefulness, confirmation, and satisfaction in the classical continued use theory for nonmedical information systems positively affect continuance intention. We also found that the path and mechanism of users' feelings regarding autonomy, competence, and relatedness generated during interactions with different gamification elements promote the continued use of mHealth apps.

(J Med Internet Res 2021;23(8):e24546) doi:10.2196/24546

KEYWORDS

mHealth app; continued use; continuance intention; gamification; self-determination theory (SDT); expectation confirmation model of information system continuance (ECM-ISC); PLS-SEM

Introduction

Background

Global health continues to face significant challenges, and health management is an important research direction. In 2016, 41 million people died of noncommunicable diseases globally, equivalent to 71% of global deaths [1]. One study showed that after 1 year of health management intervention, subjects’ risks of hypertension and hyperglycemia were reduced by 42.78% and 31.13%, respectively [2]. Moreover, every dollar invested in health management can reduce the cost of medical care by US $1 to $3 [3]. With mobile health (mHealth) technology maturity, the traditional health management model has gradually shifted to a management model based on mobile health.

With the popularization of smartphones, mHealth centered around mHealth apps has become an important means of health management. Related research is also increasing every year, mainly focusing on the health management effect, framework design, and user behavior regarding mHealth apps. According to estimates, more than 250,000 mHealth apps were on the market in 2016. It is estimated that there will be 4.68 billion mobile phone users worldwide in 2019, and the number of mHealth apps will achieve an annual growth rate of 41% between 2015 and 2020 [4]. mHealth can provide users with health intervention [5] and effective health evaluation indicators [6] and improve the communication between patients and health care professionals [7]. The number of studies on mHealth apps is also increasing every year. As of July 6, 2020, there were 688 articles related to mHealth apps in PubMed. Most studies have explored the intervention effect of mHealth apps [8-11]. Some studies focus on the development and design of mHealth apps to increase personalization and efficiency in health management services [12]. A few studies analyzed the acceptance and continued use of mHealth apps [13,14].

The effectiveness of health management based on an mHealth app is affected by user behavior. Short-term use can hardly achieve the expected goal of health management. However, the continued use of an mHealth app is not optimistic; most users only use it 4 times, and 25% use it only once after installation [14]. Therefore, it is essential to explore the influencing factors on users’ continued use of mHealth apps for the sustainable development of mHealth. Previous studies have explored the factors influencing the continued use of different types of apps. The technology acceptance model, expectation confirmation theory, unified theory of acceptance and use of technology, and other theories have been used to explore the influencing factors of continued app use, such as mobile instant messaging [15], mobile social apps [16], mobile social tourism and shopping [17], and paid mobile apps [18]. Compared to other apps, mHealth apps have different application scenarios and purposes, so it is unknown whether the research conclusions in other fields apply to mHealth apps. Therefore, some studies have explored the influencing factors of the continued mHealth app use from the aspects of quality, usability, user perception, and social impact. With the increasing number of new elements incorporated into applications, gamification is widely used in mHealth apps because it is believed to have the potential to promote continued use [19]. Out of 1000 mHealth apps, 772 contain at least 1 gamification element [20]. However, the effect of gamification in different apps is diverse [21], so whether gamification in mHealth apps has a positive impact on continued use needs to be explored.

Studies on the influence of gamification on mHealth apps’ continued use are limited. Previous studies have shown that gamification has a significant impact on the intention to continue using other types of apps [20,22,23], but whether it has the same effect on mHealth apps needs to be verified. Therefore, studies are gradually trying to explain the impact of gamification on behavioral intention in mHealth apps. For example, based on motivational feedback, some studies analyzed the factors influencing the continued use of fitness apps—an important type of mHealth app in terms of effective, informational, and social feedback—and proved the effect of gamification on affective feedback [22]. Other studies analyzed the effects of utilitarian benefits, hedonic benefits, and social benefits of gamified fitness apps on users' continuance intention and expanded the analysis of multiple impacts of social benefits on users' psychology and cognition [24]. Although previous research has substantially increased the understanding of the relationship between gamification and continued use, there are some limitations. Gamification does not cause all individuals to have the same subjective feelings and encourage them to continue using mHealth apps [25]. There is less discussion about the impact of different types of gamification on users' diverse feelings, and the resulting explanation of the different effects of gamification is insufficient. Therefore, we propose the following research questions:

1. What are the impact factors on continuance intention for gamified mHealth apps? What is the relationship between the impact factors?
2. By what path does users’ feeling induced by gamification affect users’ continuance intention toward mHealth apps?

To answer these questions, we built a research model to explain how users’ feelings induced by gamification affect the continuance intention for mHealth apps based on the expectation confirmation model of information system continuance (ECM-ISC) and self-determination theory (SDT). Specifically, SDT is used to analyze the impact of different types of gamification elements on users' feelings, and different feelings are set as constructs added to the model to explore the impact mechanism and path of gamification on the continued mHealth app use.

Significance of This Study

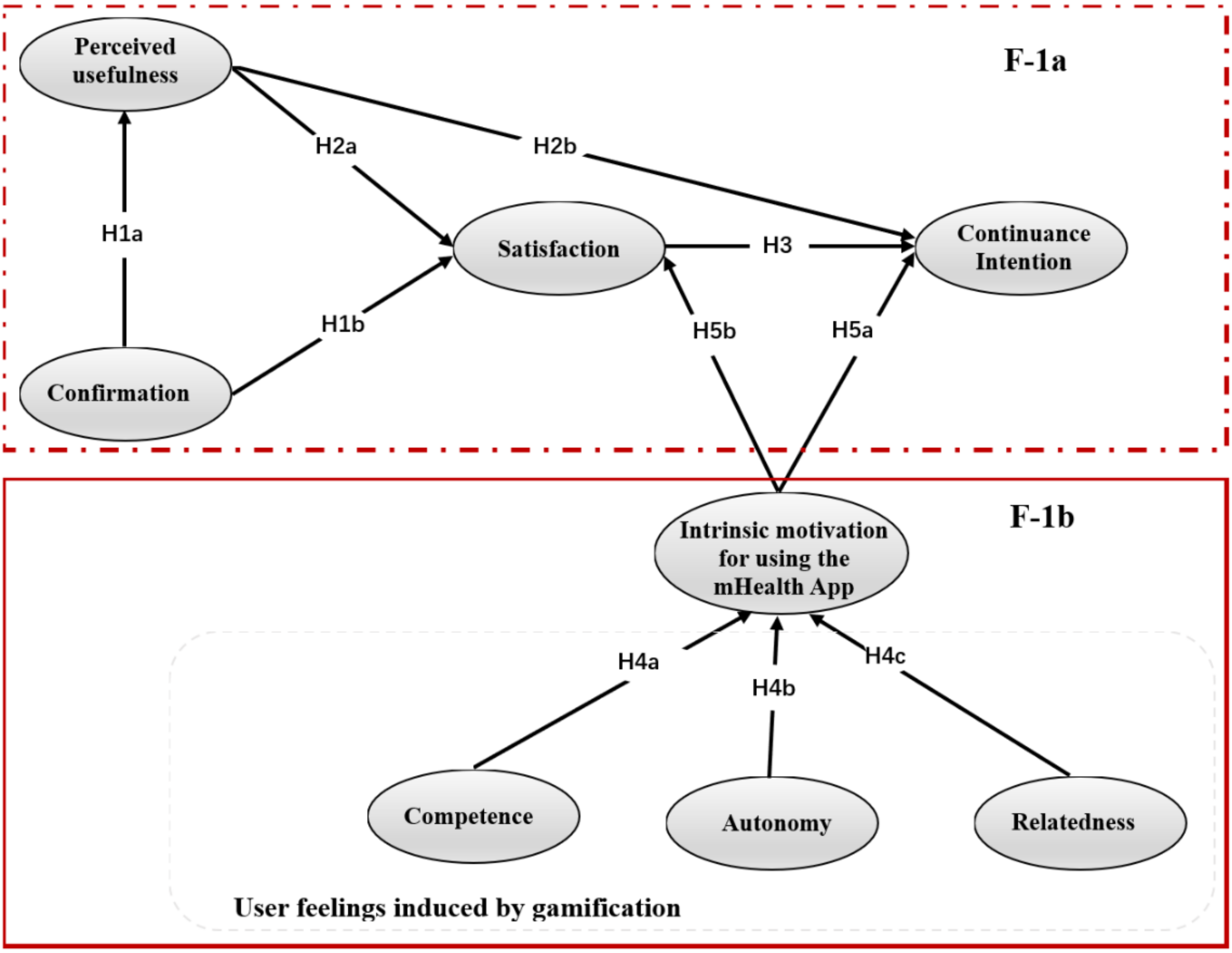
Based on the ECM-ISC model, we introduced SDT theory and innovatively used this composite model to explain psychological aspects of gamification elements that affect the motivation to use mHealth apps to promote the continuance intention of mHealth app use. Previous studies on continuance mainly focused on information systems such as hospital information systems and massive open online courses that focused on improving efficiency. There is insufficient research on the factors influencing the continued use induced by mHealth app gamification. Second, we linked gamification with the continued use model through SDT and added the users' feelings caused by different gamification elements into the research model. This paper proposes a model to explain the impact of users' feelings toward gamification on continued mHealth app use and explores the impact mechanism and path of gamification, an emerging information technology method combined with behavior change theory, on the continued mHealth app use.

We first conducted an empirical test on the continuance model of the gamified mHealth app and, for the first time, verified the applicability of the classic model ECM-ISC in the mHealth app field for Chinese users. Compared with other studies on continuance, we did not choose a specific app for verification and included more apps to increase the applicability of the results. Second, we analyzed the impact of users' feelings induced by gamification on promoting continued mHealth app use and provided a reference for mHealth app designers and researchers to promote users' continuance intention for the mHealth app and improve the effect of users' health management.

Research Model and Hypotheses

Based on the ECM-ISC, combined with SDT, we explored the impact factors of the continued use of the gamified mHealth app. The research model and assumptions are shown in Figure 1. F-1a aims to verify the applicability of ECM-ISC in the mHealth app, and F-1b explores the impact factors of feelings induced by gamification based on SDT.

Figure 1. Research model.



Hypotheses in ECM-ISC

The ECM-ISC (Figure 1, see F-1a) explains the process by which users generate continuance intention. Referring to the study by Bhattacherjee [26], the definitions of each construct are as follows: (1) confirmation is defined as users' perception of the congruence between expectations of mHealth apps use and its actual performance, (2) perceived usefulness refers to users' perception of the expected benefits of mHealth app use, (3) satisfaction is users' affect (feelings about) regarding prior mHealth apps use, (4) continuance intention is the user's willingness and tendency to continue using the mHealth app. Previous studies have revealed the relationship between these factors in nonmedical information systems [26]. Therefore, we believe that the same following assumptions exist in the mHealth app:

1. H1a: Confirmation positively influences perceived usefulness.
2. H1b: Confirmation positively influences satisfaction.
3. H2a: Perceived usefulness influences satisfaction.
4. H2b: Perceived usefulness influences continuance intention.
5. H3: Satisfaction influences continuance intention.

Hypothesis of Constructs Generated by Gamification Elements

SDT is a theory about human motivation and personality that highlights the importance of human-evolved inner resources for personality development and behavioral self-regulation. SDT defines 3 innate psychological needs: the needs for competence, autonomy, and relatedness [27]. SDT asserts that external things allow users to have autonomy, competence, and relatedness by satisfying innate psychological needs. The 3 feelings increase humans’ intrinsic motivation [28]. Gamification is the application of game design elements in nongame contexts, so the essence of gamification in mHealth app is a design method of app. The gamification element is its manifestation, and it is objective within the app. The direct stimulus of behavior motivation is the user's personality and feelings rather than external factors. Therefore, we need to explore the effect of gamification on users' feelings and then learn its impacts on behavior motivation. To achieve this goal, we chose 5 of the most common gamification elements as research objects. According to the characteristics of the different elements, this paper analyzed the impact of gamification on users' feelings. In the context of this research, autonomy refers to gamification providing users with the right to choose and set specific gamification contents independently and to determine their health management content and set goals; therefore, users have autonomy over their behavior [28]. Health management is the primary goal of the mHealth app, so we mainly explore the user's right to choose and self-controlled behaviors in health management goals.

The common achievement or progression-oriented gamification elements include points, medals, and leaderboards (PML) [21]. Therefore, we assume that the user's competence will be affected after experiencing the above gamification elements. In the context of this article, competence means that users obtain gamified feedback after completing app health management tasks, and this feedback allows uses to understand their health management capabilities. Furthermore, social networking is a social-oriented gamification element, which can leverage the functions of friending, commenting, and sharing experiences [21]. In mHealth apps, the above functions are generally utilized by the health community. This process reduces a sense of isolation from others and satisfies user needs for relatedness, making users feel relatedness [28]. Furthermore, according to SDT, competence, autonomy, and relatedness can increase the intrinsic motivation to use the mHealth app, described as follows:

1. H4a: Competence induced by gamification positively influences intrinsic motivation for using the mHealth app.
2. H4b: Autonomy induced by gamification positively influences intrinsic motivation for using the mHealth app.
3. H4c: Relatedness induced by gamification positively influences intrinsic motivation for using the mHealth app.

Relationship Between Intrinsic Motivation for Using the mHealth app and Satisfaction and Continuance Intention

Intrinsic motivation drives human behaviors, emphasizing that humans carry out certain activities based on their personalities and feelings [27]. Feelings are not easy to change, so intrinsic motivation is persistent; therefore, the intrinsic motivations of the mHealth app promote continuance intention. Under the influence of their feelings and interests, users will actively use the mHealth app, resulting in satisfactory evaluations. Therefore, we hypothesize the following:

1. H5a: Intrinsic motivation for using the mHealth app positively influences continuance intention.
2. H5b: Intrinsic motivation for using the mHealth app positively influences satisfaction.

Methods

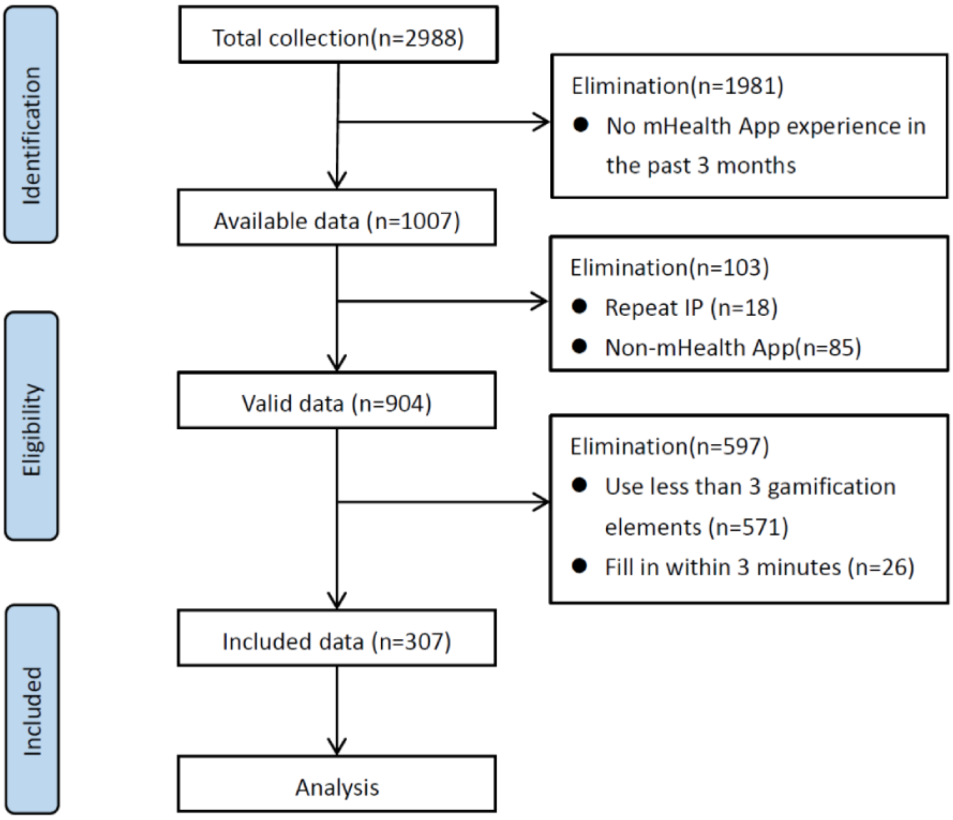
Measurement Instrument

This study constructed a new theoretical model to explore how users’ feelings induced by gamification promote the continued use of mHealth apps. It was necessary to select appropriate observable variables for the constructs to verify our model. To develop the measurement instrument of the survey, we adapted previously validated scales to our research context. To design the measurement instrument reasonably, we adjusted the previously validated scales according to the experimental purpose. For the feelings caused by gamification elements, items for competence were drawn from McAuley et al [29]. Items for autonomy were sourced from several studies [29-31]. Items for relatedness were from 3 studies [30-32]. Items for intrinsic motivation for using mHealth apps were from Lin [33]. Scales for ECM-ISC, items for perceived usefulness, satisfaction, confirmation, and continuance intention were drawn from validated scales [26,34,35]. We used the 7-point Likert scale to measure the items with anchors ranging from 1 (strongly disagree) to 7 (strongly agree). We used the initial questionnaire to presurvey 17 testers and modified the scale based on evaluations and suggestions to form the final questionnaire (see Multimedia Appendix 1). We used the data collection service from China's largest online survey platform to manage the survey [36].

Research Target and Data Collection

The study subjects have used an mHealth app in the past 3 months. From January 20 to February 23, 2020, we used WeChat to conduct a snowball sampling survey, posting the link to the questionnaire on the WeChat moments and WeChat group. A total of 2988 questionnaires were returned, mainly distributed across 29 regions of China. In addition, 11 responses from Chinese users were also received from the United States, Australia, and Austria. We excluded responses that included the following characteristics: no mHealth app experience in the past 3 months (n=1981, 66.30%), repeat internet protocols (IP) (n=18, 0.60%), non-mHealth app (n=85, 2.84%), use of fewer than 3 gamification elements (n=571, 19.11%), and completion within 3 minutes (n=26, 0.87%). Finally, we obtained a valid questionnaire (n=307, 10.27%). No mHealth app experience means that users replied that they had not used the mHealth app in the past 3 months. A non-mHealth app means the app used by the user was not a mHealth app. We determined whether the app is an mHealth app according to the classification system of Apple’s App Store. If the app used by the user was not included in the “health and fitness” and “medical” categories, it was regarded as a “non-mHealth app.” The data inclusion and exclusion process are shown in Figure 2. In addition, although we selected users who have used at least 3 gamification elements, some users still needed to answer a few items that include a gamification element that users did not use. For the items involving in the gamification elements that participants did not use, we input point 4 (uncertainty) on the 7-point Likert scale. This study was approved by the Biomedical Ethics Committee of Peking University, and the subjects were aware of the study’s purpose and process.

Figure 2. Data inclusion and exclusion process. IP: internet protocol; mHealth: mobile health.



Data Analysis

The partial least squares structural equation model (PLS-SEM) is more suitable for exploratory research than the covariance-based structural equation model [37]. PLS-SEM can deal with non-normal data and complex models with more latent variables [38] and has lower sample requirements [10]. Our research is exploratory, and the data are not strictly normally distributed; therefore, this study uses PLS-SEM analysis with SmartPLS (version 3.2.8; SmartPLS GmbH).

Results

Demographic Information

The demographic information for the included samples is shown in Table 1. We divided the apps into 4 categories according to their functions, including menstruation and pregnancy management, fitness and diet management, online consultation, etc. We also investigated the users' health status during follow-up to explore whether health status has a regulatory effect on continued use. The users’ health was indirectly measured by perceived health using a 7-point Likert scale from 1 (very unhealthy) to 7 (very healthy), with an average of 5.35 (SD 1.193).

Table 1. Demographic information.

Characteristics	Users, n (%)
Gender	
Male	174 (56.7)
Female	133 (43.3
Age	
≤30	101 (32.9)
31-45	132 (43.0)
46-60	67 (21.8)
≥61	7 (2.3)
Education	
High school	4 (1.3)
Junior college	22 (7.2)
College	130 (42.3)
Master’s degree and above	151 (49.2)
Income	
≤2500	60 (19.5)
2501-5000	28 (9.1)
5001-8000	57 (18.6)
8001-30,000	125 (40.7)
≥30,001	37 (12.1)
Frequency	
≤1 time per week	83 (27.0)
2-3 times per week	95 (30.9)
4-5 times per week	58 (18.9)
6-7 times per week	45 (14.7)
>7 times per week	26 (8.5)
Occupation	
Students	58 (18.9)
Medicine-related personnel	70 (22.8)
Public servants and clerks	77 (25.1)
Commercial and service personnel	28 (9.1)
Professional technical personnel	66 (21.5)
Manual workers	4 (1.3)
Unemployed and other personnel	4 (1.3)
Type of mHealth app	
Menstruation and pregnancy management	11 (3.6)
Fitness and diet management	212 (69.1)
Online consultation	37 (12.1)
Others	47 (15.3)

Analysis Result of Hypothesized Model

Measurement Model

The reliability and validity of the questionnaire were measured by confirmatory factor analysis. The results are shown in Table 2. To test the model's reliability, we calculated Cronbach's alpha and composite reliability based on the score of each item. A value of 0.5 represents acceptable reliability, and 0.7 represents good reliability [39]. To verify the convergent validity, we calculated the project load and the value of average extracted variance (AVE). The larger the factor loading, the greater the influence of the observed variable on the latent variable. The commonly adopted acceptance standard is 0.7. AVE represents the degree of aggregation of different items in the same construct, and a value greater than 0.5 represents acceptable convergence validity [39]. As shown in Table 2, all the values of Cronbach's alpha and composite reliabilities are above 0.8, the AVE of each construct is above 0.6, and the loading weights for each item are above 0.7, indicating good reliability for all constructs and good convergent validity. As shown in Table 3, we compare the correlation between the AVE square root of each construct and other constructs to show that the observed variable of the same construct is not related to other latent variables. The results show that the square root of AVE of each structure is greater than the interconstruct correlations, showing good discriminant validity [40]. Therefore, we conclude that the results of the measurement model are sufficient to test the hypotheses in the model.

Considering that all constructs are measured by a scale, common method bias needs to be tested to eliminate systematic errors caused by the same data source and measurement environment. We performed a statistical analysis that is suitable for the PLS-SEM model. We added a common factor to the model, which contains all items, and calculated each indicator's substantive variances explained by the principal construct and by the method [41]. The average substantive variance of the indicators is 0.904, while the average method-based variance is 0.007. The ratio of substantive variance to method variance is approximately 129:1 (see Multimedia Appendix 2). Given the small magnitude and insignificance of method variance, we believe that there is no serious common method bias in this study. The multicollinearity was checked by calculating the variance inflation factor (VIF) [42]. The VIF values of all the regressions were all substantially below the cutoff value of 10, which means there are no severe multicollinearity problems (see Multimedia Appendix 2).

Table 2. Construct reliability and convergent validity.

Construct and items	Factor loadings	Composite reliability	Average variance extracted	Cronbach alphas
Autonomy (AUT)				
AUT1	0.908	0.935	0.827	0.896
AUT2	0.914			
AUT3	0.906			
Competence (COMP)				
COMP1	0.806	0.891	0.6732	0.817
COMP2	0.874			
COMP3	0.883			
Relatedness (REL)				
REL1	0.847	0.936	0.786	0.909
REL2	0.864			
REL3	0.925			
REL4	0.907			
Continuance intention (CI)				
CI1	0.888	0.925	0.754	0.891
CI2	0.799			
CI3	0.884			
CI4	0.900			
Confirmation (CONF)				
CONF1	0.943	0.967	0.906	0.948
CONF2	0.968			
CONF3	0.945			
Intrinsic motivation for using the mHealth app (MOT)				
MOT1	0.914	0.949	0.823	0.928
MOT2	0.926			
MOT3	0.926			
MOT4	0.862			
Perceived usefulness (USE)				
USE1	0.931	0.961	0.892	0.940
USE2	0.961			
USE3	0.941			
Satisfaction (SAT)				
SAT1	0.932	0.949	0.861	0.919
SAT2	0.916			
SAT3	0.937			

Table 3. Discriminant validity.

	AUT[a]	COMP[b]	CONF[c]	CI[d]	MOT[e]	USE[f]	REL[g]	SAT[h]
AUT	*0.909* [i]							
COMP	0.355	*0.855*						
CONF	0.445	0.490	*0.952*					
CI	0.427	0.458	0.779	*0.869*				
MOT	0.464	0.546	0.760	0.710	*0.907*			
USE	0.505	0.545	0.859	0.754	0.785	*0.945*		
REL	0.182	0.547	0.352	0.310	0.411	0.362	*0.886*	
SAT	0.512	0.507	0.859	0.781	0.820	0.849	0.373	*0.928*

[a] AUT: autonomy.

[b] COMP: competence.

[c] CONF: confirmation.

[d] CI: continuance intention.

[e] MOT: intrinsic motivation for using the mHealth app.

[f] USE: perceived usefulness.

[g] REL: relatedness.

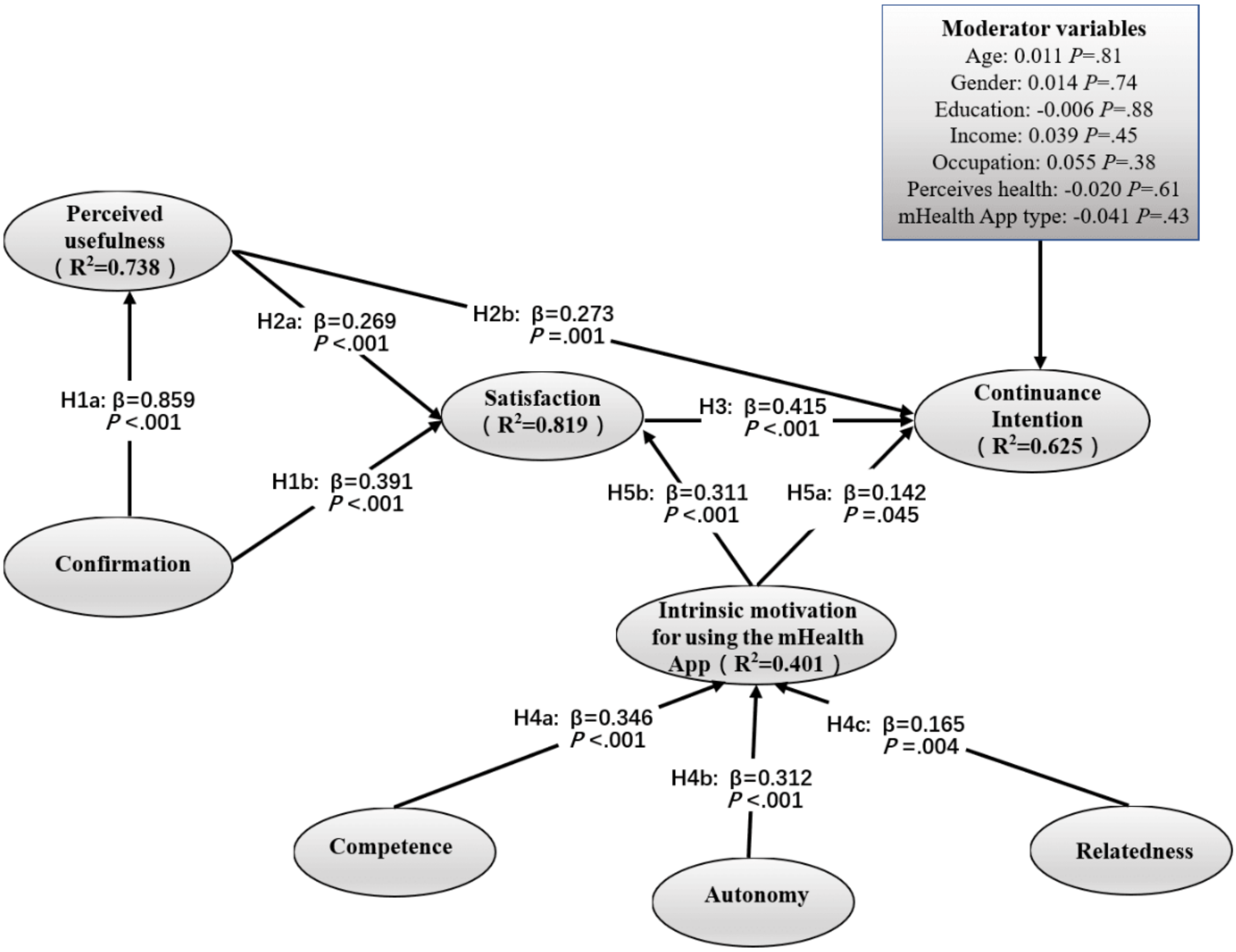
[h] SAT: satisfaction.

[i] Italics refers to the square roots of average variance extracted.

Structural Model

Through the bootstrapping analysis in PLS-SEM [43], we tested the hypotheses in the research model. As shown in Figure 3, there is an explanation for variance R^2 and standardized path coefficient β. R^2 refers to the variance of endogenous latent variables that can be explained by exogenous latent variables. The exogenous latent variable explains the variance, and β represents the correlation between the variables. First, regarding the applicability of ECM-ISC in an mHealth app, confirmation significantly affected both perceived usefulness and satisfaction, and perceived usefulness significantly affected satisfaction and continuance intention for the mHealth app. In addition, satisfaction significantly impacted the continuance intention for the mHealth app. Therefore, H1a, H1b, H2a, H2b, and H3 were supported. The above results tested the applicability of ECM-ISC in the mHealth app. Second, the constructs related to gamification, competence, autonomy, and relatedness significantly influenced the intrinsic motivation for using an mHealth app. Therefore, H4a, H4b, and H4c were all supported. Intrinsic motivation for using an mHealth app significantly impacted satisfaction and continuance intention for the mHealth app; thus, H5a and H5b were supported. As shown in Figure 3, age, gender, education, income, occupation, and other social demographic characteristics have no significant moderating effects on the results.

Figure 3. Analysis results of structural model. mHealth: mobile health.


Moderator variables
Age: 0.011 P=.81
Gender: 0.014 P=.74
Education: -0.006 P=.88
Income: 0.039 P=.45
Occupation: 0.055 P=.38
Perceives health: -0.020 P=.61
mHealth App type: -0.041 P=.43
Perceived usefulness (R²=0.738)
H2a: β=0.269 P<.001
H2b: β=0.273 P=.001
H1a: β=0.859 P<.001
Satisfaction (R²=0.819)
H3: β=0.415 P<.001
Continuance Intention (R²=0.625)
H1b: β=0.391 P<.001
H5b: β=0.311 P<.001
H5a: β=0.142 P=.045
Confirmation
Intrinsic motivation for using the mHealth App (R²=0.401)
H4a: β=0.346 P<.001
H4c: β=0.165 P=.004
H4b: β=0.312 P<.001
Competence
Autonomy
Relatedness

Results of Mediation Analysis

PLS-SEM and bootstrapping analysis results included direct effects, indirect effects, and total effects between variables. These allowed us to directly use variance accounted for (VAF) to analyze the mediating effects between variables. VAF measures the ratio of indirect effects to total effects. A VAF greater than 0.8 is generally considered a complete mediation effect, and a partial mediation effect is between 0.2 and 0.8 [44]. As shown in Table 4, all the effects were significant, whereas satisfaction was a partial mediator variable between motivation for using an mHealth app and continuance intention for the mHealth app and between perceived usefulness and continuance intention for the mHealth app. In addition, perceived usefulness was a partial mediator variable between confirmation and satisfaction.

Table 4. Mediation analysis.

Independent variable	Mediating variable	Dependent variable	Indirect effects		Total effects		VAF[a]
			β	*P* value	β	*P* value	
Intrinsic motivation	Satisfaction	Continuance	.129	<.001	.277	<.001	0.466
Usefulness	Satisfaction	Continuance	.112	.002	.386	<.001	0.290
Confirmation	Usefulness	Satisfaction	.231	<.001	.621	<.001	0.372

[a] VAR: variance accounted for.

Discussion

Most Current Gamification in the mHealth App Is Unattractive to Users

In the study, 1007 users had mHealth app experience in the past 3 months, but 248 (24.63%) users believed they had not used gamification elements. Auditing the questionnaires of the 248 users, we found that 109 (43.95%) of them used apps that contained at least 1 gamification element. Gamification has not attracted the users' attention and interest because some gamification advocates believe that the golden combination of PML can work in any app [25], leading some designers to simply combine gamification elements applied in the mHealth app with no characteristics and without considering the users' unique needs and intrinsic motivations. Although gamification elements are widely used in mHealth apps, the design process lacks an understanding of gamification and users' feelings; therefore, they cannot attract users.

The Impact Path of Gamification on Continuance Intention for mHealth Apps

This study shows that users' autonomy, competence, and relatedness interacting with different gamification elements positively promote their motivation for using an mHealth app, followed by the continuance intention for the mHealth app. This path also explains the ineffectiveness of gamification on some users. For example, if users do not feel autonomy, competence, and relatedness in the process of interacting with gamification, then gamification will not have a significant impact on the continued use of an mHealth app for this type of user.

The explanation of the impact path is as follows. First, users hope to make personalized settings in the mHealth app, such as health management tasks and user image [45]. The personalized settings allow users to experience tailor-made services, and they cater to users' interests. Autonomy is generated during the process of experiencing gamification elements such as optional goals, which increases intrinsic motivation for using the mHealth app. Second, social interaction within the mHealth app promotes the users' intention for use, indicating that the more online friends and the more frequent the interactions between users, the more significant the impact of social interaction on users' mHealth app use [46,47]. Users share health management experiences through online communities. In experiencing gamification elements with social orientation, there are interactive behaviors among users. These behaviors help users receive support and care and feel relatedness, increasing the intrinsic motivation for using the mHealth app. Finally, this study confirmed that competence caused by gamification promoted motivation for using the mHealth app. The representative gamification elements that affect competence in this study are PML. In mHealth apps, users need to complete health management tasks (eg, exercise, diet control, etc) to obtain such elements. These elements show users' progress in health management, enhancing their competence and increasing their motivation to use mHealth apps. However, it is difficult for users to change their health status after completing a health management task, and long-term and frequent use can achieve the desired effect [14]. Therefore, it is very important to reasonably design gamification elements to induce users' sense of competence, relatedness, and autonomy, increasing users' motivation to use mHealth apps.

The Influence Mechanism of Gamification on the Continuance Intention of the mHealth App

In the continued use model of the gamified mHealth app, significant impacts include confirmation, perceived usefulness, satisfaction, motivation for using the mHealth app, autonomy, competence, relatedness, and continuance intention. Among them, satisfaction and motivation for using the mHealth app have direct effects on continuance intention. Satisfaction mediates the influence of motivation for using the mHealth app on continuance intention.

High intrinsic motivation means that users' interests and feelings, which are not easily affected by external factors, are satisfied. Therefore, intrinsic motivation increases satisfaction. Furthermore, it is more likely that the user's inner emotional motivation will continue to exist, so the intrinsic motivation for using the mHealth app promotes continuance intention. This is consistent with conclusions by Cruz et al [48] and and Chang [49]. The results of the mediating effect analysis showed that satisfaction is the mediating factor explaining continuance intention, which is consistent with the conclusions from Hsiao et al [16]. Therefore, we concluded that the motivation for using an mHealth app has both direct and indirect effects on mHealth app continuance intention.

Limitations and Future Research

First, we collected data using snowball sampling, and the questionnaire was distributed through WeChat. Although the participants were distributed across the country, the sample was not enough to represent all mHealth app users, which may cause bias in the results.

Second, although we explored the impact of 5 common gamification elements on users' feelings, we did not analyze the specific role of each element and the differences between the impact of each gamification element on users' feelings. However, research in this direction is of great significance for understanding gamification. Therefore, our subsequent research focus will distinguish the effects of different types of gamification elements on users' feelings and continuance intention.

Conclusion

Based on ECM-ISC and SDT, we established a theoretical model of gamified mHealth app continuance and validated the proposed research model. Specifically, this study not only verifies the applicability of ECM-ISC to mHealth apps but also confirms that perceived usefulness, confirmation, and satisfaction positively impact mHealth app continuance intention. In terms of the impact mechanism, the competence users experience in their interaction with achievement and progress-oriented gamification elements in mHealth application, the relatedness users experience during their interaction with social-oriented gamification elements, and the autonomy users experience in their interaction with personalized service gamification elements enhance the intrinsic motivation of using mHealth apps, thus promoting the continuance intention.

Acknowledgments

This study was sponsored by the National Natural Science Foundation of China (grants 81771937 and 81871455) and PKU-Baidu Fund (grants 2019BD008). It was partially supported by the Luzhou Science and Technology Bureau project, constructing a health big data platform and service system for the elderly care institutions integrated with medical care, and the Interdisciplinary Research Funding Program for Doctoral Students of Jilin University (101832020DJX081).

Authors' Contributions

TW performed the experiments and analyzed the data. TW, LF, and XZ contributed reagents, materials, or analysis tools and prepared the figures and tables. WW and J Lei conceived and designed the experiments. J Lei authored or reviewed drafts of the paper. J Liang, KA, and MJ made revisions to the manuscript. All authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Measurement instrument.
[DOCX File , 24 KB - jmir_v23i8e24546_app1.docx]

Multimedia Appendix 2
CMV and Multicollinearity analysis.
[DOCX File , 20 KB - jmir_v23i8e24546_app2.docx]

Multimedia Appendix 3
Total collection’s demographic information.
[DOCX File , 17 KB - jmir_v23i8e24546_app3.docx]

References

1. World health statistics 2020: monitoring health for the SDGs, sustainable development goals. World Health Organization. 2020. URL: https://apps.who.int/iris/bitstream/handle/10665/332070/9789240005105-eng.pdf [accessed 2020-07-08]
2. Loeppke R, Edington DW, Bég S. Impact of the prevention plan on employee health risk reduction. Popul Health Manag 2010 Oct;13(5):275-284 [FREE Full text] [doi: 10.1089/pop.2010.0027] [Medline: 20879909]
3. Study: Wellness Programs Saved $1 to $3 per Dollar Spent. URL: https://www.shrm.org/resourcesandtools/hr-topics/benefits/pages/wellness-dollars-saved.aspx [accessed 2020-07-08]
4. Liew MS, Zhang J, See J, Ong YL. Usability Challenges for Health and Wellness Mobile Apps: Mixed-Methods Study Among mHealth Experts and Consumers. JMIR Mhealth Uhealth 2019 Jan 30;7(1):e12160 [FREE Full text] [doi: 10.2196/12160] [Medline: 30698528]
5. Kao C, Liebovitz DM. Consumer Mobile Health Apps: Current State, Barriers, and Future Directions. PM R 2017 May;9(5S):S106-S115. [doi: 10.1016/j.pmrj.2017.02.018] [Medline: 28527495]
6. Xie J, Wen D, Liang L, Jia Y, Gao L, Lei J. Evaluating the Validity of Current Mainstream Wearable Devices in Fitness Tracking Under Various Physical Activities: Comparative Study. JMIR Mhealth Uhealth 2018 Apr 12;6(4):e94 [FREE Full text] [doi: 10.2196/mhealth.9754] [Medline: 29650506]
7. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The Impact of mHealth Interventions: Systematic Review of Systematic Reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
8. Buis LR, Dawood K, Kadri R, Dawood R, Richardson CR, Djuric Z, et al. Improving Blood Pressure Among African Americans With Hypertension Using a Mobile Health Approach (the MI-BP App): Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2019 Jan 25;8(1):e12601 [FREE Full text] [doi: 10.2196/12601] [Medline: 30681965]
9. Kosse RC, Bouvy ML, Belitser SV, de Vries TW, van der Wal PS, Koster ES. Effective Engagement of Adolescent Asthma Patients With Mobile Health-Supporting Medication Adherence. JMIR Mhealth Uhealth 2019 Mar 27;7(3):e12411 [FREE Full text] [doi: 10.2196/12411] [Medline: 30916664]
10. Doocy S, Paik KE, Lyles E, Hei TH, Fahed Z, Winkler E, et al. Guidelines and mHealth to Improve Quality of Hypertension and Type 2 Diabetes Care for Vulnerable Populations in Lebanon: Longitudinal Cohort Study. JMIR Mhealth Uhealth 2017 Oct 18;5(10):e158 [FREE Full text] [doi: 10.2196/mhealth.7745] [Medline: 29046266]
11. Jiang Y, West BT, Barton DL, Harris MR. Acceptance and Use of eHealth/mHealth Applications for Self-Management Among Cancer Survivors. Stud Health Technol Inform 2017;245:131-135 [FREE Full text] [Medline: 29295067]
12. Klee P, Bussien C, Castellsague M, Combescure C, Dirlewanger M, Girardin C, et al. An Intervention by a Patient-Designed Do-It-Yourself Mobile Device App Reduces HbA1c in Children and Adolescents with Type 1 Diabetes: A Randomized Double-Crossover Study. Diabetes Technol Ther 2018 Dec;20(12):797-805. [doi: 10.1089/dia.2018.0255] [Medline: 30403495]

13. Morano JP, Clauson K, Zhou Z, Escobar-Viera CG, Lieb S, Chen IK, et al. Attitudes, Beliefs, and Willingness Toward the Use of mHealth Tools for Medication Adherence in the Florida mHealth Adherence Project for People Living With HIV (FL-mAPP): Pilot Questionnaire Study. JMIR Mhealth Uhealth 2019 Jul 03;7(7):e12900 [FREE Full text] [doi: 10.2196/12900] [Medline: 31271150]
14. Vaghefi I, Tulu B. The Continued Use of Mobile Health Apps: Insights From a Longitudinal Study. JMIR Mhealth Uhealth 2019 Aug 29;7(8):e12983 [FREE Full text] [doi: 10.2196/12983] [Medline: 31469081]
15. Oghuma AP, Libaque-Saenz CF, Wong SF, Chang Y. An expectation-confirmation model of continuance intention to use mobile instant messaging. Telematics and Informatics 2016 Feb;33(1):34-47. [doi: 10.1016/j.tele.2015.05.006]
16. Hsiao CH, Chang JJ, Tang KY. Exploring the influential factors in continuance usage of mobile social Apps: Satisfaction, habit, and customer value perspectives. Telematics and Informatics 2016 May;33(2):342-355. [doi: 10.1016/j.tele.2015.08.014]
17. Hew JJ, Leong LY, Tan GWH, Lee VH, Ooi KB. Mobile social tourism shopping: A dual-stage analysis of a multi-mediation model. Tourism Management 2018 Jun;66:121-139. [doi: 10.1016/j.tourman.2017.10.005]
18. Hsu CL, Lin JCC. What drives purchase intention for paid mobile apps? – An expectation confirmation model with perceived value. Electronic Commerce Research and Applications 2015 Jan;14(1):46-57. [doi: 10.1016/j.elerap.2014.11.003]
19. Ibarra-Herrera CC, Carrizosa A, Yunes-Rojas JA, Mata-Gómez MA. Design of an app based on gamification and storytelling as a tool for biology courses. Int J Interact Des Manuf 2019 Jun 13;13(4):1271-1282. [doi: 10.1007/s12008-019-00600-8]
20. Zhang Y, Xiao S, Zhou G. User continuance of a green behavior mobile application in China: An empirical study of Ant Forest. Journal of Cleaner Production 2020 Jan;242:118497. [doi: 10.1016/j.jclepro.2019.118497]
21. Koivisto J, Hamari J. The rise of motivational information systems: A review of gamification research. International Journal of Information Management 2019 Apr;45:191-210. [doi: 10.1016/j.ijinfomgt.2018.10.013]
22. Hassan L, Dias A, Hamari J. How motivational feedback increases user's benefits and continued use: A study on gamification, quantified-self and social networking. International Journal of Information Management 2019 Jun;46:151-162. [doi: 10.1016/j.ijinfomgt.2018.12.004]
23. Lee C, Lee K, Lee D. Mobile Healthcare Applications and Gamification for Sustained Health Maintenance. Sustainability 2017 May 08;9(5):772. [doi: 10.3390/su9050772]
24. Hamari J, Koivisto J. Why do people use gamification services? International Journal of Information Management 2015 Aug;35(4):419-431. [doi: 10.1016/j.ijinfomgt.2015.04.006]
25. Cugelman B. Gamification: what it is and why it matters to digital health behavior change developers. JMIR Serious Games 2013;1(1):e3 [FREE Full text] [doi: 10.2196/games.3139] [Medline: 25658754]
26. Bhattacherjee A. Understanding Information Systems Continuance: An Expectation-Confirmation Model. MIS Quarterly 2001 Sep;25(3):351. [doi: 10.2307/3250921]
27. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist 2000;55(1):68-78. [doi: 10.1037/0003-066X.55.1.68]
28. Ryan, Deci. Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. Contemp Educ Psychol 2000 Jan;25(1):54-67. [doi: 10.1006/ceps.1999.1020] [Medline: 10620381]
29. McAuley E, Duncan T, Tammen VV. Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: a confirmatory factor analysis. Res Q Exerc Sport 1989 Mar;60(1):48-58. [doi: 10.1080/02701367.1989.10607413] [Medline: 2489825]
30. Inchamnan W. An Analysis of Creative Process Learning in Computer Game Activities Through Player Experiences. IAFOR J Educ 2016 Sep;4(2):119-139. [doi: 10.22492/ije.4.2.07]
31. La Guardia JG, Ryan RM, Couchman CE, Deci EL. Within-person variation in security of attachment: A self-determination theory perspective on attachment, need fulfillment, and well-being. Journal of Personality and Social Psychology 2000;79(3):367-384. [doi: 10.1037/0022-3514.79.3.367]
32. Lee Y, Lee J, Hwang Y. Relating motivation to information and communication technology acceptance: Self-determination theory perspective. Computers in Human Behavior 2015 Oct;51:418-428. [doi: 10.1016/j.chb.2015.05.021]
33. Lin CP, Tsai YH, Chiu CK. Modeling Customer Loyalty from an Integrative Perspective of Self-Determination Theory and Expectation–Confirmation Theory. J Bus Psychol 2009 May 1;24(3):315-326. [doi: 10.1007/s10869-009-9110-8]
34. Rahi S, Abd. Ghani M. Integration of expectation confirmation theory and self-determination theory in internet banking continuance intention. JSTPM 2019 Oct 2;10(3):533-550. [doi: 10.1108/jstpm-06-2018-0057]
35. Hsiao JL, Chen RF. Understanding Determinants of Health Care Professionals' Perspectives on Mobile Health Continuance and Performance. JMIR Med Inform 2019 Mar 18;7(1):e12350 [FREE Full text] [doi: 10.2196/12350] [Medline: 30882353]
36. Questionnaire star_Questionnaire/onlien examination. URL: https://www.wjx.cn/ [accessed 2021-07-28]
37. Wu T, Deng Z, Zhang D, Buchanan PR, Zha D, Wang R. Seeking and using intention of health information from doctors in social media: The effect of doctor-consumer interaction. Int J Med Inform 2018 Jul;115:106-113. [doi: 10.1016/j.ijmedinf.2018.04.009] [Medline: 29779712]
38. Hair JF, Ringle CM, Sarstedt M. PLS-SEM: Indeed a Silver Bullet. The Journal of Marketing Theory and Practice 2011 Mar 15;19(2):139-152. [doi: 10.2753/MTP1069-6679190202]

39. Wu T, Deng Z, Chen Z, Zhang D, Wu X, Wang R. Predictors of Patients' Loyalty Toward Doctors on Web-Based Health Communities: Cross-Sectional Study. J Med Internet Res 2019 Sep 03;21(9):e14484 [FREE Full text] [doi: 10.2196/14484] [Medline: 31482855]
40. Fornell C, Larcker DF. Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. Journal of Marketing Research 2018 Nov 28;18(1):39-50. [doi: 10.1177/002224378101800104]
41. Liang, Saraf, Hu, Xue. Assimilation of Enterprise Systems: The Effect of Institutional Pressures and the Mediating Role of Top Management. MIS Quarterly 2007;31(1):59. [doi: 10.2307/25148781]
42. Fang J, Zhao Z, Wen C, Wang R. Design and performance attributes driving mobile travel application engagement. International Journal of Information Management 2017 Aug;37(4):269-283. [doi: 10.1016/j.ijinfomgt.2017.03.003]
43. Wu T, Deng Z, Chen Z, Zhang D, Wang R, Wu X. Predictors of Patients' Intention to Interact With Doctors in Web-Based Health Communities in China: Cross-Sectional Study. J Med Internet Res 2019 Jun 07;21(6):e13693 [FREE Full text] [doi: 10.2196/13693] [Medline: 31199296]
44. Hair JF, Hult GTM, Ringle CM, Sarstedt M. A primer on partial least squares structural equation modeling (PLS-SEM), 2nd edition. Los Angeles: SAGE Publications; 2017.
45. Milward J, Deluca P, Drummond C, Watson R, Dunne J, Kimergård A. Usability Testing of the BRANCH Smartphone App Designed to Reduce Harmful Drinking in Young Adults. JMIR Mhealth Uhealth 2017 Aug 08;5(8):e109 [FREE Full text] [doi: 10.2196/mhealth.7836] [Medline: 28790022]
46. Mendiola MF, Kalnicki M, Lindenauer S. Valuable features in mobile health apps for patients and consumers: content analysis of apps and user ratings. JMIR Mhealth Uhealth 2015 May 13;3(2):e40 [FREE Full text] [doi: 10.2196/mhealth.4283] [Medline: 25972309]
47. Hamari J, Koivisto J. “Working out for likes”: An empirical study on social influence in exercise gamification. Computers in Human Behavior 2015 Sep;50:333-347. [doi: 10.1016/j.chb.2015.04.018]
48. Cruz P, Salo J, Gallego PM, Laukkanen T. Heavy users of e-banking and Customer Experience Management: evidences on intrinsic motivation. IJEB 2010;8(2):187. [doi: 10.1504/ijeb.2010.032093]
49. Chang CC, Liang CY, Yan CF, Tseng JS. The Impact of College Students’ Intrinsic and Extrinsic Motivation on Continuance Intention to Use English Mobile Learning Systems. Asia-Pacific Edu Res 2013 May;22(2):181-192. [doi: 10.1007/s40299-012-0011-7]

Abbreviations

AVE: average extracted variance
ECM-ISC: expectation confirmation model of information system continuance
IP: internet protocol
mHealth: mobile health
PLS-SEM: partial least squares structural equation model
PML: points, medals, and leaderboards
SDT: self-determination theory
VAF: variance accounted for
VIF: variance inflation factor

Edited by R Kukafka; submitted 24.09.20; peer-reviewed by M Schmidt-Kraepelin, T Wu; comments to author 17.10.20; revised version received 09.12.20; accepted 13.07.21; published 12.08.21.

Please cite as:
Wang T, Fan L, Zheng X, Wang W, Liang J, An K, Ju M, Lei J
The Impact of Gamification-Induced Users' Feelings on the Continued Use of mHealth Apps: A Structural Equation Model With the Self-Determination Theory Approach
J Med Internet Res 2021;23(8):e24546
URL: https://www.jmir.org/2021/8/e24546
doi:10.2196/24546
PMID:34387550

©Tong Wang, Lingye Fan, Xu Zheng, Wei Wang, Jun Liang, Kai An, Mei Ju, Jianbo Lei. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 12.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

The Use and Effectiveness of an Online Diagnostic Support System for Blood Film Interpretation: Comparative Observational Study

Claire Hutchinson[1,2], MRCP, PhD; Michelle Brereton[3], PhD; Julie Adams[3], PhD; Barbara De La Salle[4], PhD; Jon Sims[4], BSc, FIBMS; Keith Hyde[3], PhD; Richard Chasty[5], MD; Rachel Brown[3], MRCP; Karen Rees-Unwin[6], PhD; John Burthem[3,6], PhD

[1]Medicine, Dentistry and Human Sciences, Faculty of Health, University of Plymouth, Plymouth, United Kingdom
[2]University Hospitals Plymouth NHS Trust, Plymouth, United Kingdom
[3]Manchester Foundation Trust, Manchester, United Kingdom
[4]UK NEQAS Haematology, Watford, United Kingdom
[5]The Christie NHS Foundation Trust, Manchester, United Kingdom
[6]Division of Cancer Sciences, University of Manchester, Manchester, United Kingdom

Corresponding Author:
John Burthem, PhD
Manchester Foundation Trust
Oxford Road
Manchester, M13 9PL
United Kingdom
Phone: 44 1617011909
Email: John.Burthem@mft.nhs.uk

Abstract

Background: The recognition and interpretation of abnormal blood cell morphology is often the first step in diagnosing underlying serious systemic illness or leukemia. Supporting the staff who interpret blood film morphology is therefore essential for a safe laboratory service. This paper describes an open-access, web-based decision support tool, developed by the authors to support morphological diagnosis, arising from earlier studies identifying mechanisms of error in blood film reporting. The effectiveness of this intervention was assessed using the unique resource offered by the online digital morphology Continuing Professional Development scheme (DM scheme) offered by the UK National External Quality Assessment Service for Haematology, with more than 3000 registered users. This allowed the effectiveness of decision support to be tested within a defined user group, each of whom viewed and interpreted the morphology of identical digital blood films.

Objective: The primary objective of the study was to test the effectiveness of the decision support system in supporting users to identify and interpret abnormal morphological features. The secondary objective was to determine the pattern and frequency of use of the system for different case types, and to determine how users perceived the support in terms of their confidence in decision-making.

Methods: This was a comparative study of identical blood films evaluated either with or without decision support. Selected earlier cases from the DM scheme were rereleased as new cases but with decision support made available; this allowed a comparison of data sets for identical cases with or without decision support. To address the primary objectives, the study used quantitative evaluation and statistical comparisons of the identification and interpretation of morphological features between the two different case releases. To address the secondary objective, the use of decision support was assessed using web analytical tools, while a questionnaire was used to assess user perceptions of the system.

Results: Cases evaluated with the aid of decision support had significantly improved accuracy of identification for relevant morphological features (mean improvement 9.8%) and the interpretation of those features (mean improvement 11%). The improvement was particularly significant for cases with higher complexity or for rarer diagnoses. Analysis of website usage demonstrated a high frequency of access for web pages relevant to each case (mean 9298 for each case, range 2661-24,276). Users reported that the decision support website increased their confidence for feature identification (4.8/5) and interpretation (4.3/5), both within the context of training (4.6/5) and also in their wider laboratory practice (4.4/5).

Conclusions: The findings of this study demonstrate that directed online decision support for blood morphology evaluation improves accuracy and confidence in the context of educational evaluation of digital films, with effectiveness potentially extending to wider laboratory use.

(J Med Internet Res 2021;23(8):e20815) doi:10.2196/20815

KEYWORDS

blood cell morphology; decision support; external quality assessment in hematology; diagnosis; digital morphology; morphology education

Introduction

Background

The morphological assessment of blood cells in different diseases is a “gatekeeper” investigation for abnormal findings identified using automated blood analyzers [1]. Important decisions are made following all levels of blood film review as morphologists must identify abnormal features, and also interpret their significance [2,3]. Those decisions range from the content of comments concerning the appearances of blood cells, through to urgent direct communication of important findings that might have an immediate impact on patient care [3,4]. Support for morphological decision-making is therefore vital, and traditionally this has been provided through training at the microscope, supplemented by the use of books and posters and the ability to refer decisions to experienced colleagues. However, providing good quality local support at all times has become challenging, as changing shift patterns and service models can reduce available support, and also impact on the cognitive processes required for accurate assessment [5,6]. Responses have included the increased use of computer-based support including semiautomated image recognition systems [7-9], as well as the increasing use of online information. Although internet-prepared resources can be very valuable, it is vital that they have accurate content and an accessible platform of delivery and are open access. It is also essential that the provision of diagnostic support is linked to objective assessment of the value of that support [10].

Prior Work

The authors of this paper have more than 15 years of experience in delivering online training through various media, including provision to developing countries [11,12] and through the UK National External Quality Assessment Service (UK NEQAS) Haematology Digital Morphology Continuing Professional Development scheme (DM CPD scheme) established in 2005 [13]. This digital scheme grew to meet the need for professional development in morphological diagnosis [14] and has more than 3000 registered users in the UK and internationally. The system has a good correlation with glass slide morphology, and has been well received by users [13]. In a recent study, the authors used data from this CPD scheme to address the mechanisms of error by morphologists interpreting blood film appearances, aiming to identify interventions that could improve diagnostic outcome [15]. The data from that study identified distinct patterns of error and suggested that diagnostic accuracy could be improved and mistakes could be reduced through the provision of information that specifically addressed these major sources of error.

Approach

This paper describes a web-based system for education and decision support, made available through desktop or mobile devices to support the morphological assessment of blood films. The system was developed to address the varied approaches required to assess cases of different types and complexity. In this paper, we assess the design considerations for the decision support website based on our previous work. Using the unique resources offered by the UK NEQAS Haematology DM CPD scheme, we then evaluate the user experience and effectiveness of the system in improving the outcomes of blood film assessment.

Methods

The Digital Morphology CPD Scheme

The Digital Morphology CPD scheme has been described in earlier publications [14]. In brief, more than 50 continuous microscopic fields were captured using Zeiss Axio Imager.M1 with a Zeiss Plan-Apochromat 63×/1.4 oil lens, then stitched using the photomerge function of Adobe Photoshop CS4 (version 11.02; Adobe Inc) and optimized for sharpness (unsharp mask function, 1 pixel up to 100%), lighting, and color balance (curves function). The final image was reviewed to ensure accurate and sufficient representation of the diagnostic features for each case, then uploaded to either Digital Slidebox (SlidePath, Leica Biosystems) or EQATE UK NEQAS Haematology Online (Certus Technology Associates Ltd). Each system provided equivalent functionality for image viewing and data submission (representative images from the EQATE system are shown in Multimedia Appendix 1). Users of the UK NEQAS DM scheme viewed the images via web browser, submitting their morphological selections and preferred diagnosis. Feedback was made using online forms with free-text responses or user grading (1-5; Google Forms, Google). The 5 cases that were rereleased used identical films. The cases were given new narratives and feedback available after case completion, and participants had no direct access to the previous release of these cases.

The Decision Support System

This was developed by the authors, and approved for use with the UK NEQAS DM CPD scheme by the Morphology Special Advisory Group of UK NEQAS General Haematology. Content was reviewed by the author team with feedback from users. The approach is outlined in the Results section; images used were “typical cell forms” (drawn using the range of functions of Photoshop CS4) or photographed images. Supporting written information was developed using the authors' experience, together with published sources.

The system was delivered using a MediaWiki platform (version 1.28.3; MediaWiki) to allow compatibility across a range of mobile devices and desktop systems. Settings prevent user edits. Consistent with the Wiki Foundation principles, readers do not impart any personal information. Links to the system from the CPD scheme were provided with each case, and the system does not restrict access by individual users. Page use data was obtained through host analytics (1&1 Ionos) to solely provide the date, number of page access episodes, and geographical location (country); this was not linked to any retained personally identifiable information.

Data Analysis and Statistics

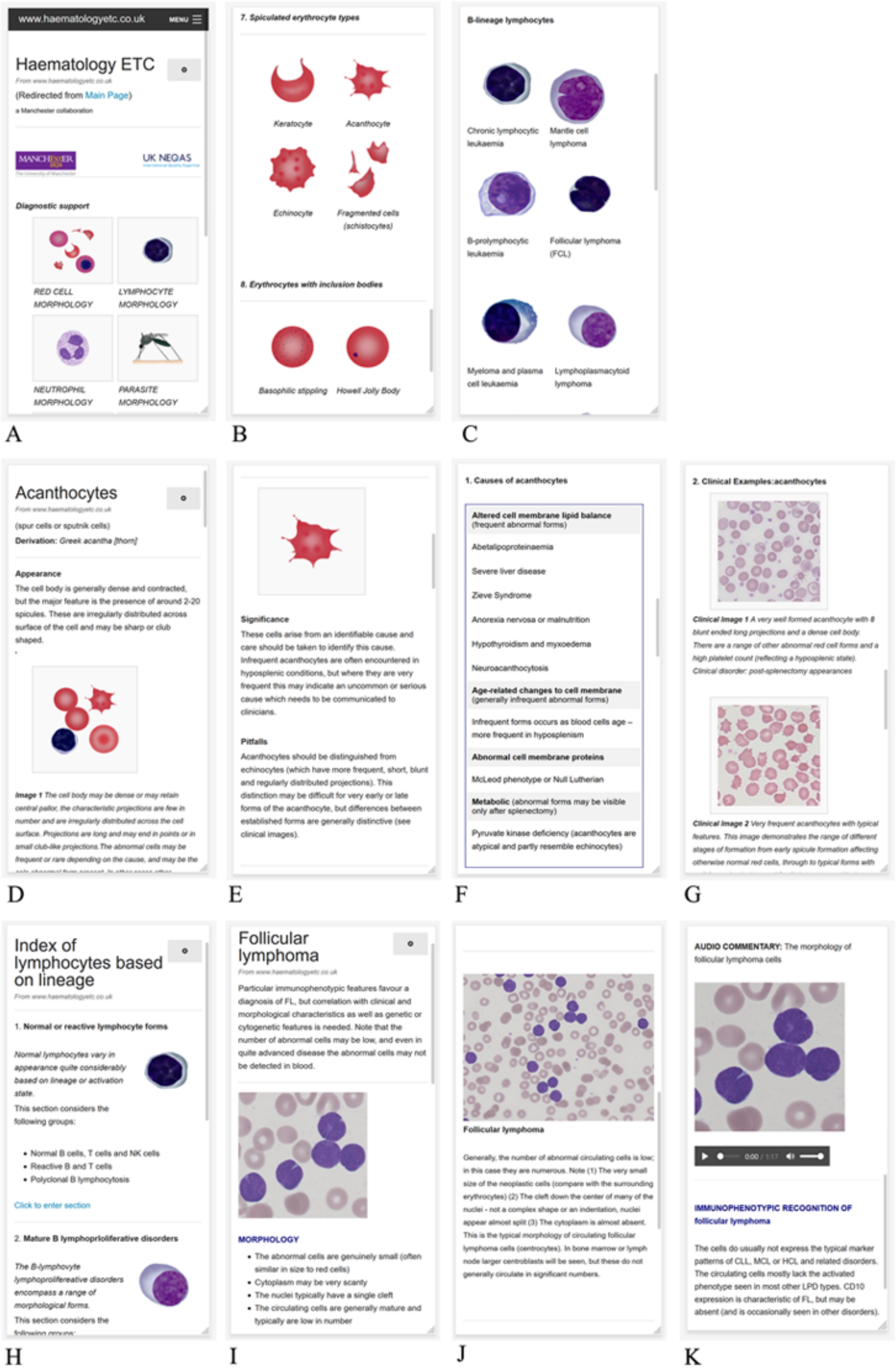
The participant responses for each DM CPD case (fully anonymized) were exported to Excel (Microsoft Corp). The data analyzed comprised the following: the number of participants, self-reported experience level, and for each film and user response, the morphological features selected as being present and the preferred diagnosis or diagnostic comment (free text). All free-text responses were then coded according to diagnosis by the authors to facilitate statistical comparison. Statistical comparison was performed using GraphPad Prism (version 7.03; GraphPad Software Inc), with details given in figure legends.

Results

Design Principles and Implementation

A decision support system was developed and made available to users through desktop or mobile devices. Of 320 participants responding to our survey, the majority used a desktop system (n=294, 92%), although laptops (31%), tablets (n=30, 9%), and mobile phones (n=30, 9%) were frequently used, either in addition to or instead of a desktop device. The system has image-driven menus that show typical morphological forms (Figure 1A-C) linked to more detailed interpretive and confirmatory support through text and clinical images. For each cell type, information was presented that reflects the different requirements of interpretation: red blood cell sections centered particularly on identifying the range of different appearances that may occur together on a blood film, with supporting information focusing on the significance of the combinations of features found (Figure 1D-G). The white blood cell sections focused predominantly on the evaluation of cytological features of individual cells (Figure 1H-K). The full working decision support site has free open access [16].

Figure 1. Representative screenshot details from the decision support app. Panels A-C show details of image-driven indexing: (A) main index, (B) red blood cell index, and (C) lymphocyte index. Panels D-G show details from red blood cell pages (acanthocyte morphology page) showing the following: (D) detailed cell description, (E) significance and pitfalls in diagnosis, (F) causes, and (G) clinical examples. Panels H-K show details from lymphocyte pages: (H) functional index, (I) follicular lymphoma main description, (J) clinical context, and (K) audio commentary.


www.haematologyetc.co.uk
MENU
Haematology ETC
From www.haematologyetc.co.uk
(Redirected from Main Page)
a Manchester collaboration
MANCHESTER 1824
UK NEQAS
Diagnostic support
RED CELL MORPHOLOGY
LYMPHOCYTE MORPHOLOGY
NEUTROPHIL MORPHOLOGY
PARASITE MORPHOLOGY
A
7. Spiculated erythrocyte types
Keratocyte
Acanthocyte
Echinocyte
Fragmented cells (schistocytes)
8. Erythrocytes with inclusion bodies
Basophilic stippling
Howell Jolly Body
B
B-lineage lymphocytes
Chronic lymphocytic leukaemia
Mantle cell lymphoma
B-prolymphocytic leukaemia
Follicular lymphoma (FCL)
Myeloma and plasma cell leukaemia
Lymphoplasmacytoid lymphoma
C
D
E
F
G
H
I
J
K
0:00 / 1:17

Acanthocytes

From www.haematologyetc.co.uk

(spur cells or sputnik cells)

Derivation: *Greek acantha [thorn]*

Appearance

The cell body is generally dense and contracted, but the major feature is the presence of around 2-20 spicules. These are irregularly distributed across surface of the cell and may be sharp or club shaped.

Image 1 *The cell body may be dense or may retain central pallor, the characteristic projections are few in number and are irregularly distributed across the cell surface. Projections are long and may end in points or in small club-like projections.The abnormal cells may be frequent or rare depending on the cause, and may be the*

Significance

These cells arise from an identifiable cause and care should be taken to identify this cause. Infrequent acanthocytes are often encountered in hyposplenic conditions, but where they are very frequent this may indicate an uncommon or serious cause which needs to be communicated to clinicians.

Pitfalls

Acanthocytes should be distinguished from echinocytes (which have more frequent, short, blunt and regularly distributed projections). This distinction may be difficult for very early or late forms of the acanthocyte, but differences between established forms are generally distinctive (see clinical images).

1. Causes of acanthocytes

Altered cell membrane lipid balance (frequent abnormal forms)

Abetalipoproteinaemia

Severe liver disease

Zieve Syndrome

Anorexia nervosa or malnutrition

Hypothyroidism and myxoedema

Neuroacanthocytosis

Age-related changes to cell membrane (generally infrequent abnormal forms)

Infrequent forms occurs as blood cells age – more frequent in hyposplenism

Abnormal cell membrane proteins

McLeod phenotype or Null Lutherian

Metabolic (abnormal forms may be visible only after splenectomy)

Pyruvate kinase deficiency (acanthocytes are atypical and partly resemble echinocytes)

2. Clinical Examples:acanthocytes

Clinical Image 1 *A very well formed acanthocyte with 8 blunt ended long projections and a dense cell body. There are a range of other abnormal red cell forms and a high platelet count (reflecting a hyposplenic state). Clinical disorder: post-splenectomy appearances*

Clinical Image 2 *Very frequent acanthocytes with typical features. This image demonstrates the range of different stages of formation from early spicule formation affecting otherwise normal red cells, through to typical forms with*

Index of lymphocytes based on lineage

From www.haematologyetc.co.uk

1. **Normal or reactive lymphocyte forms**

Normal lymphocytes vary in appearance quite considerably based on lineage or activation state.

This section considers the following groups:

- Normal B cells, T cells and NK cells
- Reactive B and T cells
- Polyclonal B lymphocytosis

Click to enter section

2. **Mature B lymphoprloliferative disorders**

The B-lymphovyte lymphoprolifereative disorders encompass a range of morphological forms.

This section considers the following groups:

Follicular lymphoma

From www.haematologyetc.co.uk

Particular immunophenotypic features favour a diagnosis of FL, but correlation with clinical and morphological characteristics as well as genetic or cytogenetic features is needed. Note that the number of abnormal cells may be low, and even in quite advanced disease the abnormal cells may not be detected in blood.

MORPHOLOGY

- The abnormal cells are genuinely small (often similar in size to red cells)
- Cytoplasm may be very scanty
- The nuclei typically have a single cleft
- The circulating cells are generally mature and typically are low in number

Follicular lymphoma

Generally, the number of abnormal circulating cells is low; in this case they are numerous. Note (1) The very small size of the neoplastic cells (compare with the surrounding erythrocytes) (2) The cleft down the center of many of the nuclei - not a complex shape or an indentation, nuclei appear almost split (3) The cytoplasm is almost absent. This is the typical morphology of circulating follicular lymphoma cells (centrocytes). In bone marrow or lymph node larger centroblasts will be seen, but these do not generally circulate in significant numbers.

AUDIO COMMENTARY: The morphology of follicular lymphoma cells

IMMUNOPHENOTYPIC RECOGNITION OF follicular lymphoma

The cells do usually not express the typical marker patterns of CLL, MCL or HCL and related disorders. The circulating cells mostly lack the activated phenotype seen in most other LPD types. CD10 expression is characteristic of FL, but may be absent (and is occasionally seen in other disorders).

The Use and Outcomes of the Decision Support System

Use of the decision support system was optional, and participants were also free to access any other reference source. To map use of the system, analytic software was used to record access to pages from each country, without recording personal identifiable data. This showed that, when compared with the reference period, there was a significant increased use of the site during the 3-week period when each case was available (Figure 2A). The average number of UK individuals accessing the main site menu each week during each reference period (when no CPD case was available) was 932, compared to a mean of 2359 per week when cases were active (mean difference 1427, *P*=.02, n=23,000 observations, Wilcoxon test of paired data). The mean number of relevant pages accessed during each case release was 9298 (range 2661-24,276). The mean number of participants who submitted entries to the CPD scheme for each case was 1022, suggesting that a substantial proportion of participants visited the site on at least one occasion and viewed multiple relevant pages. Additionally, a weak to moderate trend for increased general use of the site was also noted, with an overall trend to approximately double over 12 months when time points reflecting "live" CPD assessments were excluded (Figure 2B). When the use of decision support was compared with the degree of difficulty of the case—as indicated by the proportion of incorrect responses (% incorrect) when the case was first released—it was found that the use of decision support was very strongly correlated with the difficulty of the case (R^2=0.85; Figure 2C). Finally, users were directly asked their perception of the system regarding different aspects of use. For users responding to the survey request (n=230), most reported increased confidence in identification of features (4.6/5) and diagnosis (4.4/5). The system was valued highly in the context of education (4.7/5) and in routine laboratory usage (4.4/5; Figure 3).

Figure 2. Access to decision support and accuracy of Continuing Professional Development completion. (A) web activity over time (red line) with trend line (black line) indicating changing weekly access numbers; each new case release is indicated by an arrow. (B) The same data excluding access associated with case availability, showing an approximate doubling of access by UK users and weak to moderate correlation. (C) A high number of web visits in the most recent case release was very strongly correlated with increasing case difficulty (as indicated by the percentage of correct responses when first released; R^2=0.85). Numbers refer to individual cases as assigned in Table 1. (D) Accuracy of reporting of the major morphological features for each case in the two separate releases, showing a consistently improved accuracy when decision support was available. (E) The number of participants accurately identifying the abnormal findings was increased in the more recent release, particularly in the more difficult cases. (F) The level of improvement in outcome was strongly correlated with increased access of relevant pages of the decision support tool (numbers refer to individual cases as assigned in Table 1). Statistical testing for (D) and (E): chi-square test with Yates correction. ATLL: adult T-cell leukemia lymphoma; G6PD: glucose-6-phosphate dehydrogenase; MAHA: microangiopathic hemolytic anemia; SAO: Southeast Asian ovalocytosis.

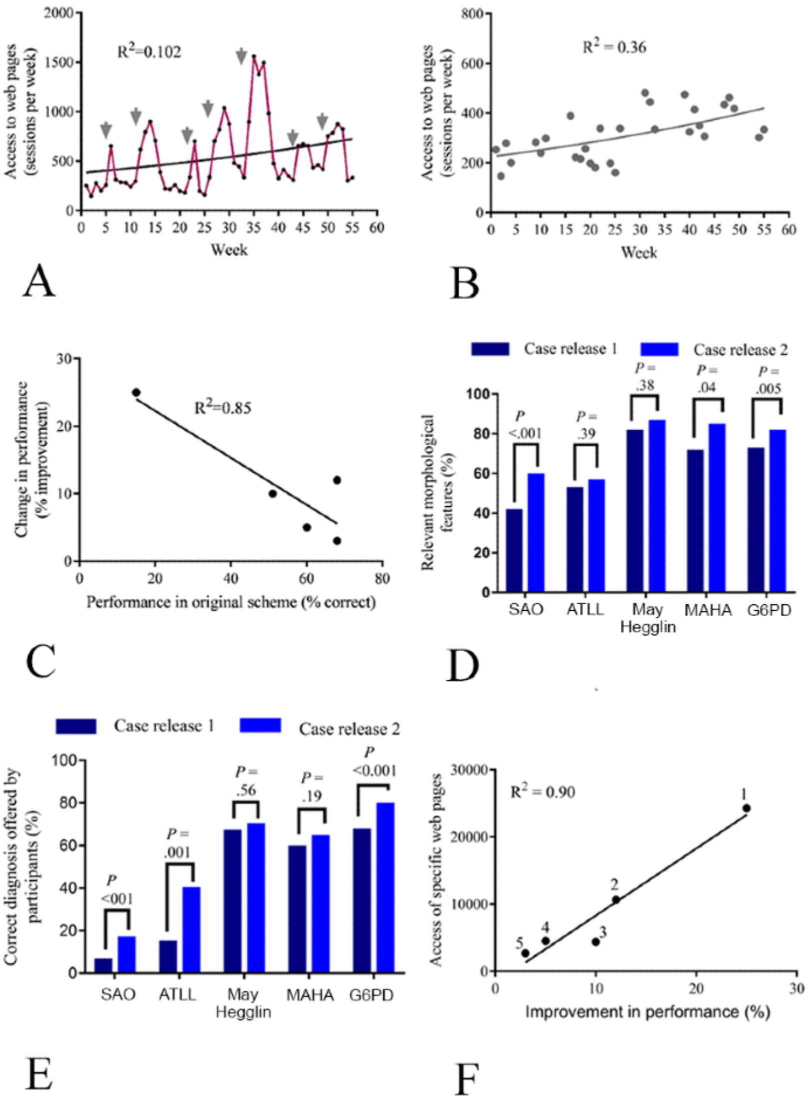
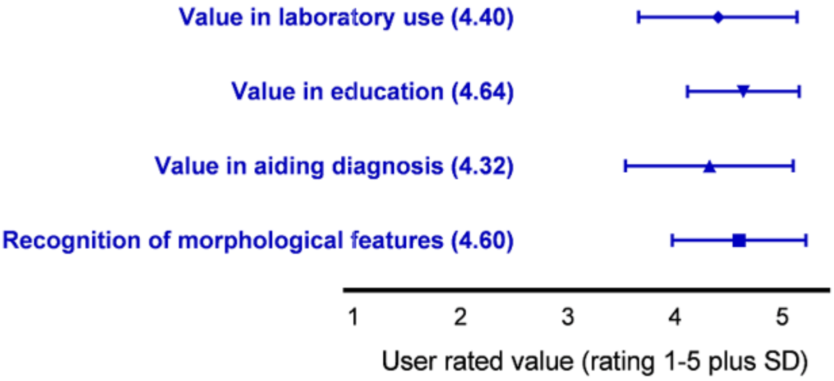


Figure 3. Participant interaction with and perception of the system. Users were asked their view of the value of the system in the contexts indicated. Users rated the system 1-5, where 1 was the lowest score and 5 was the highest score (data shows mean and SD of a user sample of 200). Users were asked the following questions: (1) Do you find HaematologyETC useful in helping you to identify morphological features in CPD cases? (2) Do you find HaematologyETC useful in helping you to reach a diagnosis in CPD cases? (3) How would you rate HaematologyETC as an educational tool? (4) Do you think HaematologyETC would be useful in assisting morphology reporting in your laboratory?



Overall, these data reflect widespread use of the decision support system to support CPD scheme activity, with selective use of those pages according to the perception of case difficulty, as well as increased general use of the system outside of the scheme. We therefore went on to assess whether the decision support tool led to improved accuracy of morphological assessment and interpretation. To achieve this, 5 digital images that had previously been used in the CPD scheme were rereleased in identical form, but now with users given access to the decision support system. To avoid these cases being recalled by participants, no case was used if it had been released in the previous 4 years, and the mean interval was 5.4 years between the two release dates (Table 1). The number of registered users was found to be increased in the more recent period and access to the cases was increased overall from 5375 to 6663 participants (24%); however, the experience of users in each assessment period was comparable (the self-reported proportion of qualified staff reporting blood films regularly for a representative case was similar between the two periods [536/713 participants, 75% compared with 976/1354, 72%]). This analysis showed that the availability of the decision support tool for participants in the more recent cases led to a more accurate recognition of the morphological forms present in the digital image (based on those features selected by participants as most characteristic of the blood smear). The improvement varied between cases, with a mean of 9.8% (n=2902 users in initial 5 case releases and n=3922 in recent 5 case releases, range of improvement 4%-18%, *P*=.02, Student *t* test of matched pairs; Figure 2D). Furthermore, the diagnostic accuracy was also consistently improved in the more recent cases when decision support was available, with an overall improvement of 11% (n=2902 users in initial 5 case releases and n=3922 in recent 5 case releases, range of improvement 3%-25%, *P*=.02, Wilcoxon test of paired *t* test data), and the greatest improvement seen in “more difficult” cases where diagnostic accuracy in the initial case release was lower (Figure 2E). These more difficult cases were also associated with much higher rates of use of the decision support tool (as reflected by web access of pages relevant to the case diagnosis; Figure 2F).

Overall, analysis showed that the accuracy of diagnosis was significantly better across the full case set (total participants contributing diagnoses n=6824, mean accuracy 55% versus 44%, *P*=.02, Student *t* test for paired data).

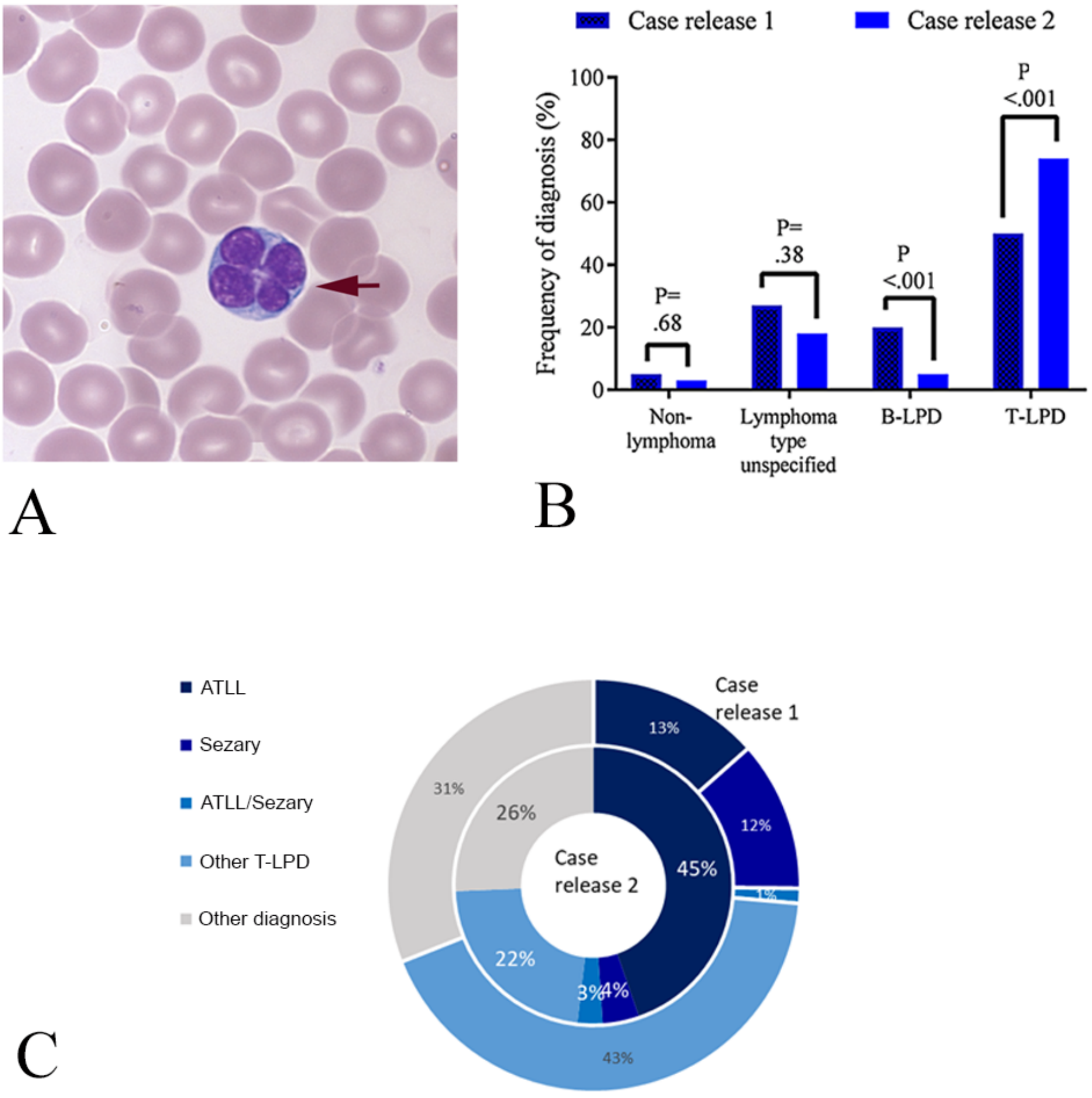
Table 1. Case number, description, and participant data.

Case number	Diagnosis	First release date	Participants, first release, n	Participants, second release, n	Change, %
1	Southeast Asian ovalocytosis with Epstein-Barr virus infection (Southeast Asian ovalocytosis/glandular fever)	2014	991	1392	40
2	Adult T-cell leukemia lymphoma (ATLL)	2015	1102	1418	28
3	May-Hegglin anomaly (May-Hegglin)	2013	1104	1271	15
4	Microangiopathic hemolytic anemia (MAHA)	2015	1277	1544	21
5	Oxidative hemolysis in glucose-6-phosphate dehydrogenase deficiency (G6PD)	2011	901	1092	21

Effect on Decision-making in Complex Morphological Cases

The case of adult T-cell leukemia lymphoma (ATLL) demonstrated the challenge of detailed morphological evaluation and interpretation of abnormalities when there were small numbers of abnormal cells (Figure 4A). The case required participants first to recognize that the lymphoid cell population was abnormal, then to make a detailed morphological assessment of the cells present, and ascribe a likely diagnosis based on that assessment. This case was difficult for participants when first released, with only around 16% making the correct diagnosis. When rereleased, the case attracted particularly frequent use of the decision support website, with more than 18,000 relevant pages accessed by participants, of which more than 3700 were specific for the diagnosis of ATLL. This high use of support was associated with an improved accuracy in the different stages of the diagnostic process; in the initial release, 829 participants answered the case, with 1282 participants in the more recent release. Although participants identified the lymphocytosis with equal frequency, those in the more recent case release group were more likely to identify the cells as neoplastic (n=615, 48% versus n=328, 40%, *P*<.001, chi-square test, Fisher exact test) and were also significantly more likely to suggest that the cells represented a neoplastic T-cell disorder (934/1282 participants versus 414/829, 73% versus 50%; *P*<.001, chi-square test with Yates correction; Figure 4B). This was associated with a significantly greater likelihood of ascribing a precise diagnosis to the case (rather than a generic diagnosis of T-cell lymphoproliferative disorder) as well as a much higher recognition that the diagnosis was ATLL (513/1282, 45% versus 108/829, 13%; *P*<.001, chi-square test with Yates correction; Figure 4C). These findings suggest that support for the detailed cytological evaluation of rare cell types was widely used, and that this resulted in improved accuracy in the morphological evaluation of uncommon disorders.

Figure 4. Detailed analysis of the case with ATLL. (A) Typical morphological features of the case, demonstrating a particularly characteristic large abnormal lymphocyte showing a lobulated "clover leaf" nucleus (arrowed). (B) A greater proportion of participants identified the diagnosis as lymphoma; of these, a significantly greater proportion identified the T-cell nature of the disease. (C) Pie chart comparing diagnostic choice for white blood cell features in release 1 and release 2, showing significantly improved rates of specific diagnosis and of diagnostic accuracy in the more recent release (release 2). Statistical testing for (B): chi-square test with Yates correction. ATLL: adult T-cell leukemia lymphoma.

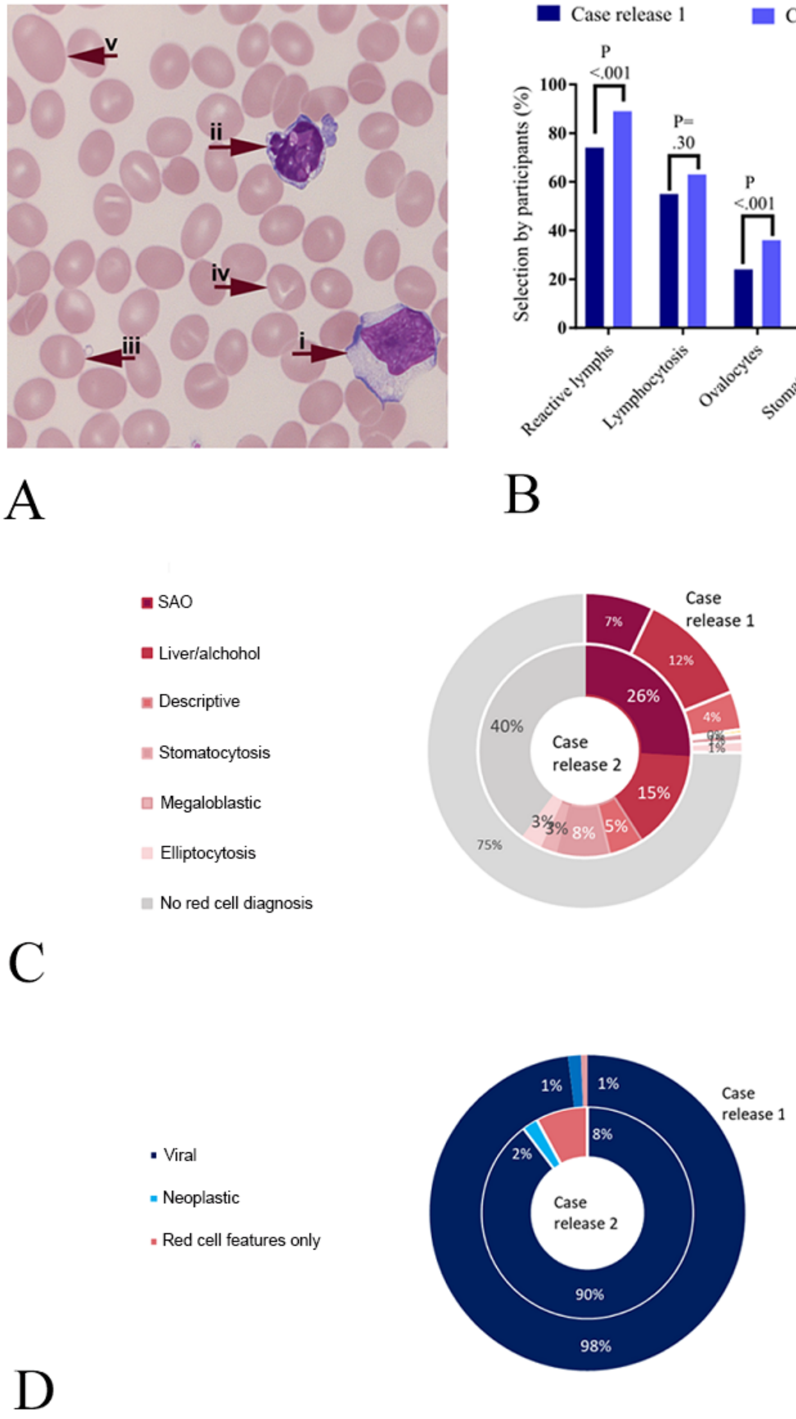


Effect on Decision-making in the Presence of Multiple Abnormal Features

For those diagnoses where a range of abnormal morphological features are present, interpretation requires strategies that depend on the identification, classification, and prioritization of the abnormal features. The accuracy of these strategies depends very much on the skill and knowledge of the morphologist, and may result in unhelpful simplification processes such as elimination bias (where features are ignored or assigned inappropriately low significance for diagnosis) or framing bias (where features are considered to fit a single preferred diagnosis, irrespective of whether this is the most likely interpretation).

Consistent with this, the most complex case in this set combined the red blood cell disorder Southeast Asian ovalocytosis (SAO) with an accompanying Epstein-Barr virus infection, leading to distinctive but complex morphology (Figure 5A). This case proved difficult for participants when first released, with a tendency for errors related to simplification strategies; in the initial release, the marked red blood cell abnormalities were all reported relatively infrequently, particularly the diagnostically important stomatocytes (Figure 5B), and red blood cell diagnoses were offered by only 245 of 991 (25%) participants. The correct interpretation of SAO was made by just 69 of 991 participants (7%) in the initial release (Figure 5C). In contrast, the features of viral illness were well recognized (reported and correctly interpreted by 961/991 participants, 97%; Figure 5D). When the case was rereleased, the decision support tool was heavily used, with more than 11,000 views of pages relevant to the case, and this was associated with a more accurate identification of abnormal morphological features (Figure 5B), with the overall consideration of red blood cell disorders in the final diagnosis increased to 841/1402 participants (60%; *P*<.001; Figure 5C) and the correct diagnosis of SAO given by 365/1402 participants, (26%; *P*<.001, chi-square test with Yates correction). Interestingly, the increased awareness of the red blood cell features led to indications that some participants used a framing bias, attempting to unify the white blood cell and red blood cell features into a single diagnosis of liver disease or malignancy (Figure 5D). A small proportion showed elimination bias, as they simply omitted any white blood cell diagnosis at all. Overall, the reduced rate of diagnosis of viral disorder was not significant (*P*=.19, chi-square test with Yates correction).

Figure 5. Detailed analysis of the case with SAO and EBV infection. (A) Typical morphological features of the case, demonstrating (i) a large activated lymphocyte, (ii) an apoptotic lymphocyte, and (iii) the erythroid abnormalities including stomatocytes, some of which have (iv) complex stoma and (v) very large oval cells. (B) Accuracy of morphological evaluation by participants, showing a higher percent selection of all relevant morphological features of the case, particularly the reporting of stomatocytes. (C) Pie chart comparing diagnostic choice for red blood cell features in release 1 and release 2, showing improved accuracy in the second release. (D) Pie chart comparing diagnostic choice for white blood cell features in release 1 and release 2, showing a trend for better accuracy in the first release. Statistical testing for (B): chi-square test with Yates correction. EBV: Epstein-Barr virus; SAO: Southeast Asian ovalocytosis.



Discussion

Principal Results

In this study, digital blood films provided within the online UK NEQAS Haematology DM CPD scheme were used to evaluate the effectiveness of a new decision support system developed by the authors to support morphological assessment in hematology. This analysis revealed that use of the decision support site was associated with a significantly improved accuracy of reporting for both morphological feature identification and the interpretation of those features. That information, together with web page analytic data and direct user feedback, has shown that the system was widely used by CPD participants and improved the performance and confidence of users.

The detailed analysis of the two most difficult cases was particularly useful to evaluate the effectiveness of support. For white blood cell types, the system focused on providing detailed help in assessing cytology of individual cell types to support recognition of characteristic features that allow discrimination between different cell forms. This help was very highly used in the assessment of the uncommon lymphoproliferative disorder ATLL, and this use was associated with a very substantial improvement both in recognizing the unusual (T-cell) nature of the disorder and in suggesting the correct diagnosis (ATLL). For red blood cell diagnoses, the aim of the system was (as for white blood cells) to provide help in recognizing abnormal forms, and additionally focused on identifying pitfalls in cell recognition and showing how combinations of features could be used to make a diagnosis. In the case of SAO, the participants more frequently reported those relevant red blood cell features present and also showed considerable improvement in suggesting the correct diagnosis (SAO).

Possible Limitations of the Study

The decision support was evaluated using an online education system and compared present users examining DM CPD cases with the historical performance of users assessing the same case. We recognize that the study design therefore contains potentially confounding factors related to the separate time points. In particular, users in the current period were likely to have increased familiarity with technology and online resources and were more likely to have access to better technical resources, screen quality, and download speed [17,18]. The study partly addresses these limitations, showing that irrespective of other factors the decision support pages were frequently accessed at times when CPD cases were available for assessment. In addition, the web pages accessed corresponded closely to the subject of the case, consistent with directed use of the system by participants. Furthermore, there was evidence of selective use of pages according to case difficulty, with a marked correlation between the number of times relevant pages were accessed and the difficulty of the case (as indicated by performance in the earlier period).

It is also worth noting that the system does not address all sources of error. The case of SAO also had reactive (“viral”) lymphocytes, which were very widely reported in the earlier release; interestingly, the improved awareness of red blood cell features was associated with a small reduction in those reporting the likely viral disorder. These errors were similar to the “heuristic errors” observed in our previous study of cases with a highly complex group of features [15] and may reflect familiar sources of bias in complex cases, such as “anchoring” or “elimination” bias, where morphologists focused only on red blood cell features.

Context and Prior Work

Providing effective support for morphological diagnosis requires an understanding of the process; the complexity and number of cells present on a blood film means that (consciously or unconsciously) morphologists employ approaches to reduce the complexity of the evaluation to allow a timely conclusion [19,20]. Briefly, the process has two phases. The first phase is “perception,” a process where the morphologist perceives whether the observed appearances differ from an expectation of normality; this is a rapid and largely subconscious process [21], whose effectiveness depends significantly on the experience of the observer [22]. The second phase employs the skills of “recollection and analysis” to evaluate particular findings in detail [23], requiring the active application of knowledge through techniques that simplify analysis, such as classification (allowing abnormal cells to be considered as classes or groups rather than as separate entities), prioritization (to focus on those elements that are considered most important to diagnosis), and elimination (to remove from consideration those elements considered unimportant to diagnosis). These techniques all require an appropriate knowledge base for accurate outcomes; when incorrectly performed, the process can lead to errors of various types (reviewed in [15,24,25]).

The support system in this paper describes an approach to providing effective “bench side” diagnostic support for the analytical phase of blood film evaluation by providing a rapid and accessible information source available at the time of evaluation focusing on the central processes of cell recognition, classification, prioritization, and interpretation. A digital platform was selected based on the advantages offered by such a system in the context of morphology. First, digital platforms enable the uploading and displaying of many images, as well as the ability to select and magnify features or to drive links or visual menu systems. Second, digital platforms are accessible to all users and are accessed through computers or mobile devices that can be freely available at the point of need. Third, this digital approach offers a flexibility that can be iteratively modified to suit user needs or changing information; this platform can be linked to other resources and websites that provide supporting or detailed information. It is important to reflect that global health resources should offer accessibility across geographical barriers; in this case, the adoption of a MediaWiki platform and free access to all users facilitates wide use. It is recognized though that language barriers often remain. However, while the decision support system developed by the authors is written solely using English language, there are positive features for future development. First, the use of a MediaWiki platform allows access to the tools developed to facilitate easier page translation for Wikipedia and related applications. These tools support potential future collaborative development of the system in different language formats.

Second, the use of image-driven menus within the system was intended to support easier navigation for those not fully familiar with morphological descriptions, but this may also enable the use of the system by a wider international group.

Conclusions

In this paper, we have demonstrated the effectiveness of this specifically designed online tool to improve the performance of morphologists in the setting of an online CPD system. However, the broader question is whether there is an expanded role for its use in general diagnosis within laboratories. The internet is gaining wider penetration in all areas of society; the public is increasingly using and familiar with internet resources, often accessed by mobile devices. Online diagnosis support can be used either at the microscope or as an adjuvant to computer-assisted diagnostic systems. The flexibility of a web-based approach may also be extended, particularly to support fast-evolving areas such as molecular diagnosis. Finally, the system may have particular value in developing countries where internet access via mobile devices may become a major point of access for teaching, training, or other services, and where low-cost flexible support can provide wider benefit if problems of adoption can be overcome [12,26].

Acknowledgments

The authors would like to thank Celgene (UK) and Sysmex (UK) for support of this work through educational grants.

Authors' Contributions

CH and MB contributed equally to the manuscript, and both took images, analyzed data, and wrote the paper. JS and BDLS provided and analyzed data and edited the manuscript. RC, JA, and RB wrote text and reviewed the paper. KH contributed to the design of the study. KRU wrote and edited the paper. JB designed the study, took images, analyzed data, and wrote the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The EQATE system. (A) The menu system in the left panel allows users to select information or provide responses to the case; the panel can be expanded by users as required to provide a larger view of the image, which can be expanded or contracted to allow higher magnification viewing. (B) The full image field in the right panel shows around 50 images taken at high magnification using an oil immersion lens allowing maximum resolution of detail, while the left panel shows the narrative that guides users through the major features of the film following completion of the case. (C) Image viewed at high power (equivalent to magnification at the time of image acquisition). (D) Additional features of the case are shown at highest magnification with an accompanying annotation following completion of the case.

[PNG File , 805 KB - jmir_v23i8e20815_app1.png]

References

1. Comar SR, Malvezzi M, Pasquini R. Evaluation of criteria of manual blood smear review following automated complete blood counts in a large university hospital. Rev Bras Hematol Hemoter 2017;39(4):306-317 [FREE Full text] [doi: 10.1016/j.bjhh.2017.06.007] [Medline: 29150102]
2. Adewoyin AS, Nwogoh B. Peripheral blood film - a review. Ann Ib Postgrad Med 2014 Dec;12(2):71-79 [FREE Full text] [Medline: 25960697]
3. Palmer L, Briggs C, McFadden S, Zini G, Burthem J, Rozenberg G, et al. ICSH recommendations for the standardization of nomenclature and grading of peripheral blood cell morphological features. Int J Lab Hematol 2015 Jun;37(3):287-303. [doi: 10.1111/ijlh.12327] [Medline: 25728865]
4. Buoro S, Da Rin G, Fanelli A, Lippi G. Harmonization of interpretative comments in laboratory hematology reporting: the recommendations of the Working Group on Diagnostic Hematology of the Italian Society of Clinical Chemistry and Clinical Molecular Biology (WGDH-SIBioC). Clin Chem Lab Med 2018 Dec 19;57(1):66-77. [doi: 10.1515/cclm-2017-0972] [Medline: 29389663]
5. Caruso C. Negative impacts of shiftwork and long work hours. Rehabil Nurs 2014;39(1):16-25 [FREE Full text] [doi: 10.1002/rnj.107] [Medline: 23780784]
6. Brown M, Tucker P, Rapport F, Hutchings H, Dahlgren A, Davies G, et al. The impact of shift patterns on junior doctors' perceptions of fatigue, training, work/life balance and the role of social support. Qual Saf Health Care 2010 Dec 25;19(6):e36-e36 [FREE Full text] [doi: 10.1136/qshc.2008.030734] [Medline: 21127102]
7. Criel M, Godefroid M, Deckers B, Devos H, Cauwelier B, Emmerechts J. Evaluation of the Red Blood Cell Advanced Software Application on the CellaVision DM96. Int J Lab Hematol 2016 Aug;38(4):366-374. [doi: 10.1111/ijlh.12497] [Medline: 27199257]

8. Yu H, Ok CY, Hesse A, Nordell P, Connor D, Sjostedt E, et al. Evaluation of an automated digital imaging system, Nextslide Digital Review Network, for examination of peripheral blood smears. Arch Pathol Lab Med 2012 Jun;136(6):660-667 [FREE Full text] [doi: 10.5858/arpa.2011-0285-OA] [Medline: 22646275]
9. Briggs C, Longair I, Slavik M, Thwaite K, Mills R, Thavaraja V, et al. Can automated blood film analysis replace the manual differential? An evaluation of the CellaVision DM96 automated image analysis system. Int J Lab Hematol 2009 Feb;31(1):48-60. [doi: 10.1111/j.1751-553X.2007.01002.x] [Medline: 18177438]
10. 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]
11. Hutchinson CV, Brereton ML, Burthem J. Digital imaging of haematological morphology. Clin Lab Haematol 2005 Dec;27(6):357-362. [doi: 10.1111/j.1365-2257.2005.00727.x] [Medline: 16307535]
12. Ahmed L, Seal LH, Ainley C, De la Salle B, Brereton M, Hyde K, et al. Web-Based Virtual Microscopy of Digitized Blood Slides for Malaria Diagnosis: An Effective Tool for Skills Assessment in Different Countries and Environments. J Med Internet Res 2016 Aug 11;18(8):e213 [FREE Full text] [doi: 10.2196/jmir.6027] [Medline: 27515009]
13. Burthem J, Brereton M, Ardern J, Hickman L, Seal L, Serrant A, et al. The use of digital 'virtual slides' in the quality assessment of haematological morphology: results of a pilot exercise involving UK NEQAS(H) participants. Br J Haematol 2005 Jul;130(2):293-296. [doi: 10.1111/j.1365-2141.2005.05597.x] [Medline: 16029459]
14. Brereton M, De La Salle B, Burthem J, Ardern J, Hickman L, Seal L, et al. Review of the UK NEQAS (H) digital morphology pilot scheme for continuing professional development accessed via the internet. Int J Lab Hematol 2008 Oct;30(5):365-371. [doi: 10.1111/j.1751-553X.2008.01086.x] [Medline: 18691344]
15. Brereton M, De La Salle B, Ardern J, Hyde K, Burthem J. Do We Know Why We Make Errors in Morphological Diagnosis? An Analysis of Approach and Decision-Making in Haematological Morphology. EBioMedicine 2015 Sep;2(9):1224-1234 [FREE Full text] [doi: 10.1016/j.ebiom.2015.07.020] [Medline: 26501122]
16. Haematologyetc. URL: https://haematologyetc.co.uk [accessed 2021-07-26]
17. Haag M, Igel C, Fischer MR, German Medical Education Society (GMA), Committee “Digitization – Technology-Assisted Learning and Teaching”, Joint working group “Technology-enhanced Teaching and Learning in Medicine (TeLL)” of the German Association for Medical Informatics, Biometry and Epidemiology (gmds) and the German Informatics Society (GI). Digital Teaching and Digital Medicine: A national initiative is needed. GMS J Med Educ 2018;35(3):Doc43 [FREE Full text] [doi: 10.3205/zma001189] [Medline: 30186953]
18. Thimbleby H. Technology and the future of healthcare. J Public Health Res 2013 Dec 01;2(3):e28 [FREE Full text] [doi: 10.4081/jphr.2013.e28] [Medline: 25170499]
19. Shah AK, Oppenheimer DM. Heuristics made easy: an effort-reduction framework. Psychol Bull 2008 Mar;134(2):207-222. [doi: 10.1037/0033-2909.134.2.207] [Medline: 18298269]
20. Wegwarth O, Gaissmaier W, Gigerenzer G. Smart strategies for doctors and doctors-in-training: heuristics in medicine. Med Educ 2009 Aug;43(8):721-728. [doi: 10.1111/j.1365-2923.2009.03359.x] [Medline: 19573016]
21. DiCarlo JJ, Zoccolan D, Rust NC. How does the brain solve visual object recognition? Neuron 2012 Feb 09;73(3):415-434 [FREE Full text] [doi: 10.1016/j.neuron.2012.01.010] [Medline: 22325196]
22. Bar M. A cortical mechanism for triggering top-down facilitation in visual object recognition. J Cogn Neurosci 2003 May 15;15(4):600-609. [doi: 10.1162/089892903321662976] [Medline: 12803970]
23. Henson RN, Rugg MD, Shallice T, Josephs O, Dolan RJ. Recollection and Familiarity in Recognition Memory: An Event-Related Functional Magnetic Resonance Imaging Study. J Neurosci 1999 May 15;19(10):3962-3972 [FREE Full text] [doi: 10.1523/JNEUROSCI.19-10-03962.1999] [Medline: 10234026]
24. Gunderman RB. Biases in radiologic reasoning. AJR Am J Roentgenol 2009 Mar;192(3):561-564. [doi: 10.2214/AJR.08.1220] [Medline: 19234247]
25. Blumenthal-Barby JS, Krieger H. Cognitive biases and heuristics in medical decision making: a critical review using a systematic search strategy. Med Decis Making 2015 May;35(4):539-557. [doi: 10.1177/0272989X14547740] [Medline: 25145577]
26. Maware S, Sai KOS. An investigation on e-resource utilisation among university students in a developing country: A case of Great Zimbabwe University. South African Journal of Information Management 2018;20(1):a860.

Abbreviations

ATLL: adult T-cell leukemia lymphoma
CPD: Continuing Professional Development
DM: Digital Morphology
SAO: Southeast Asian ovalocytosis
UK NEQAS: UK National External Quality Assessment Service for Haematology

Edited by R Kukafka; submitted 02.06.20; peer-reviewed by M Wondergem, S Sarbadhikari; comments to author 27.10.20; revised version received 28.10.20; accepted 19.04.21; published 09.08.21.

Please cite as:
Hutchinson C, Brereton M, Adams J, De La Salle B, Sims J, Hyde K, Chasty R, Brown R, Rees-Unwin K, Burthem J
The Use and Effectiveness of an Online Diagnostic Support System for Blood Film Interpretation: Comparative Observational Study
J Med Internet Res 2021;23(8):e20815
URL: https://www.jmir.org/2021/8/e20815
doi:10.2196/20815
PMID:34383663

©Claire Hutchinson, Michelle Brereton, Julie Adams, Barbara De La Salle, Jon Sims, Keith Hyde, Richard Chasty, Rachel Brown, Karen Rees-Unwin, John Burthem. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 09.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Development and Validation of Unplanned Extubation Prediction Models Using Intensive Care Unit Data: Retrospective, Comparative, Machine Learning Study

Sujeong Hur[1,2*], RN, MS; Ji Young Min[1*], MS; Junsang Yoo[3], RN, PhD; Kyunga Kim[1,4], PhD; Chi Ryang Chung[5,6], MD, PhD; Patricia C Dykes[7], RN, PhD; Won Chul Cha[1,8,9], MD

[1]Department of Digital Health, Samsung Advanced Institute for Health Sciences & Technology, Sungkyunkwan University, Seoul, Republic of Korea
[2]Department of Patient Experience Management, Samsung Medical Center, Seoul, Republic of Korea
[3]Department of Nursing, College of Nursing, Sahmyook University, Seoul, Republic of Korea
[4]Biomedical Statistics Center, Research Institute for Future Medicine, Samsung Medical Center, Seoul, Republic of Korea
[5]Department of Critical Care Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea
[6]Department of Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea
[7]Division of General Internal Medicine and Primary Care, Brigham and Women’s Hospital, Harvard Medical School, Boston, MA, United States
[8]Department of Emergency Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea
[9]Digital Innovation Center, Samsung Medical Center, Seoul, Republic of Korea
*these authors contributed equally

Corresponding Author:
Won Chul Cha, MD
Department of Emergency Medicine
Samsung Medical Center
Sungkyunkwan University School of Medicine
81 Irwon-ro, Gangnam-gu
Seoul, 06351
Republic of Korea
Phone: 82 2 3410 2053
Fax: 82 2 3410 0012
Email: wc.cha@samsung.com

Abstract

Background: Patient safety in the intensive care unit (ICU) is one of the most critical issues, and unplanned extubation (UE) is considered the most adverse event for patient safety. Prevention and early detection of such an event is an essential but difficult component of quality care.

Objective: This study aimed to develop and validate prediction models for UE in ICU patients using machine learning.

Methods: This study was conducted in an academic tertiary hospital in Seoul, Republic of Korea. The hospital had approximately 2000 inpatient beds and 120 ICU beds. As of January 2019, the hospital had approximately 9000 outpatients on a daily basis. The number of annual ICU admissions was approximately 10,000. We conducted a retrospective study between January 1, 2010, and December 31, 2018. A total of 6914 extubation cases were included. We developed a UE prediction model using machine learning algorithms, which included random forest (RF), logistic regression (LR), artificial neural network (ANN), and support vector machine (SVM). For evaluating the model’s performance, we used the area under the receiver operating characteristic curve (AUROC). The sensitivity, specificity, positive predictive value, negative predictive value, and F1 score were also determined for each model. For performance evaluation, we also used a calibration curve, the Brier score, and the integrated calibration index (ICI) to compare different models. The potential clinical usefulness of the best model at the best threshold was assessed through a net benefit approach using a decision curve.

Results: Among the 6914 extubation cases, 248 underwent UE. In the UE group, there were more males than females, higher use of physical restraints, and fewer surgeries. The incidence of UE was higher during the night shift as compared to the planned extubation group. The rate of reintubation within 24 hours and hospital mortality were higher in the UE group. The UE prediction algorithm was developed, and the AUROC for RF was 0.787, for LR was 0.762, for ANN was 0.763, and for SVM was 0.740.

Conclusions: We successfully developed and validated machine learning–based prediction models to predict UE in ICU patients using electronic health record data. The best AUROC was 0.787 and the sensitivity was 0.949, which was obtained using the RF algorithm. The RF model was well-calibrated, and the Brier score and ICI were 0.129 and 0.048, respectively. The proposed prediction model uses widely available variables to limit the additional workload on the clinician. Further, this evaluation suggests that the model holds potential for clinical usefulness.

(J Med Internet Res 2021;23(8):e23508) doi:10.2196/23508

KEYWORDS

intensive care unit; machine learning; mechanical ventilator; patient safety; unplanned extubation

Introduction

Patient safety in the intensive care unit (ICU) is a critical issue. Medical errors and adverse events can significantly impact patient outcomes [1]. Medical errors are a common occurrence in the ICU and airway-related accidents are the most frequent [2]. Adverse events related to airway and mechanical ventilation, such as unplanned extubation (UE), may lead to high rates of morbidity and mortality [3].

UE is a critical adverse event in the ICU, necessitating immediate action and treatment by the medical staff. In the literature, UE incidence rates range from 0.5 to 35.8 per 100 ventilated patients [4,5]. Previous studies have revealed that UE is associated with significant complications, such as airway injury, prolonged respiratory distress, aspiration, and hypoxemia [6]. Even after reintubation, UE remains associated with longer ICU stays [7] and an increased risk of ventilator-associated pneumonia [8].

Strategies to prevent UE include introducing a quality improvement program and novel devices [9,10]. However, for effective application of these tools, continuous screening and early detection is necessary. An electronic health record (EHR)-based prediction system could be an efficient and timely tool to provide continuous screening and early detection.

The wide establishment of advanced EHR systems has facilitated the development of machine learning prediction models [11]. These systems have shown substantial potential in predicting complex clinical conditions, such as sepsis, readmission, and cardiopulmonary resuscitation [12-14]. However, we were unable to find published examples of machine learning prediction models that were used for UE prediction. Therefore, the objective of this study was to develop and validate machine learning–based UE prediction models for patients in the ICU.

Methods

The Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement [15] was followed for reporting our multivariable prediction model.

Study Setting and Data Source

A single-center, retrospective study was conducted based on the EHR data of an academic tertiary hospital in Seoul, Republic of Korea. Data from January 2010 to December 2018 were extracted from the clinical data warehouse of the hospital, which contained deidentified clinical data for research. The hospital has approximately 2000 inpatient beds and 120 ICU beds. There are two types of ICUs: a medical ICU and a surgical ICU. In this study, 42 beds for the medical ICU and 70 beds for the surgical ICU were included. As of January 2019, there were approximately 9000 patients in the outpatient department and 250 patients in the emergency department on a daily basis. The number of annual ICU admissions is approximately 10,000.

Study Population

The study population included patients who underwent extubation in the ICU between January 1, 2010, and December 31, 2018. Patients under the age of 18 years and patients who had multiple extubation episodes were excluded from the study. Patients who had been on mechanical ventilation for less than 24 hours or for more than 2 weeks were also excluded: patients with short mechanical ventilation periods had been admitted to the ICU only for a short period of observation, and the ICU protocol was to perform tracheostomy on patients by 2 weeks from the intubation.

Outcome of Prediction Models

The risk prediction models used in this study had binary outcomes. They dealt with either the occurrence or absence of UE for an intubated ICU patient based on data from the last 8 hours.

Data Set

We constructed a data set containing UE risk factors based on a literature review, which included the following: Confusion Assessment Method for the ICU (CAM-ICU) [16], the Richmond Agitation-Sedation Scale (RASS) [17], the Glasgow Coma Scale (GCS), upper-limb motor power, lower-limb motor power, the use of physical restraints, and work shifts. Because intubated patients cannot be assessed through verbal response due to the presence of an artificial airway, the verbal response records in the GCS were not considered. All included variables were routinely recorded by a nurse in the critical care flow sheet in the ICUs. The patients’ baseline characteristics were also included in the data set, consisting of age, sex, whether the patient underwent surgery prior to ICU admission, intubation location, and reason for ICU admission.

We split the data sets periodically for development and validation. The data sets acquired between January 1, 2010, and December 31, 2015, were used for development sets. The data sets acquired between January 1, 2016, and December 31, 2018, were used for validation sets.

Data Preprocessing

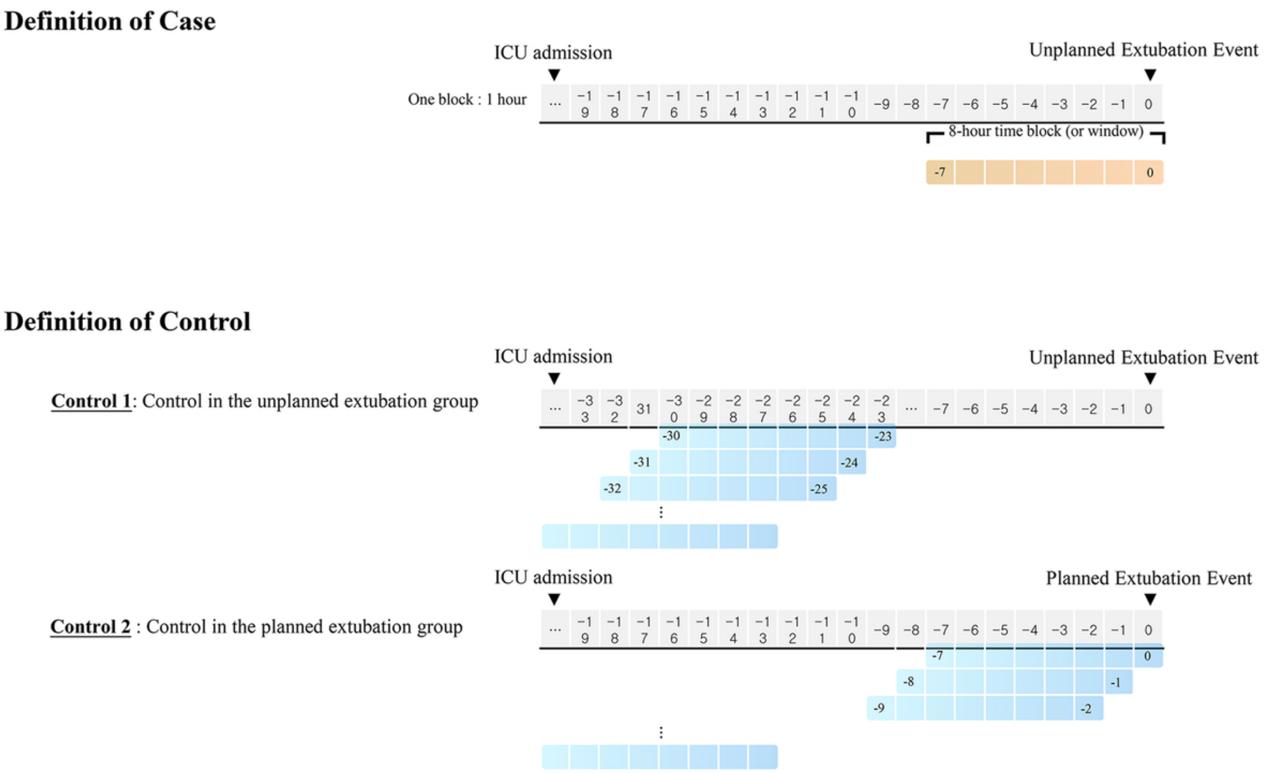
Time-Window Setting

Features related to the CAM-ICU, the RASS, the GCS, and limb motor powers changed over time in the data sets. We set up a time window to consider the changing trends over time in these time-series features. We calculated the average recording intervals for each time-series feature and set 8 hours as the size of our time window, which covered the longest interval among them; as such, we expected that at least one change for all time-series features would be considered in the 8-hour time window. In addition, the characteristics of the clinical workflow of the institution were reflected. In the ICU where the study was conducted, nurses usually worked three shifts. We considered the time point at which the change in the patient's condition could be sufficiently reflected in the EHR and, finally, an 8-hour window was selected.

Defining Cases and Controls

A moving window with an 8-hour period was used to define cases and controls. The case and control definitions using the time window in the time-series data set is shown in Figure 1. When the UE event occurred, the 8-hour time block, or window, was annotated as a case. The 8-hour time block from ICU admission to 24 hours prior to the UE event (control 1) and the 8-hour time block from ICU admission to planned extubation event (control 2) were annotated as a control.

Figure 1. Case and control definitions using the time window in the time-series data set. ICU: intensive care unit.


Definition of Case
ICU admission
Unplanned Extubation Event
One block : 1 hour
... −19 −18 −17 −16 −15 −14 −13 −12 −11 −10 −9 −8 −7 −6 −5 −4 −3 −2 −1 0
8-hour time block (or window)
-7
0
Definition of Control
ICU admission
Unplanned Extubation Event
Control 1: Control in the unplanned extubation group
... −33 −32 31 −30 −29 −28 −27 −26 −25 −24 −23 ... −7 −6 −5 −4 −3 −2 −1 0
-30
-23
-31
-24
-32
-25
⋮
ICU admission
Planned Extubation Event
Control 2 : Control in the planned extubation group
... −19 −18 −17 −16 −15 −14 −13 −12 −11 −10 −9 −8 −7 −6 −5 −4 −3 −2 −1 0
-7
0
-8
-1
-9
-2
⋮

Time-Series Feature Handling

Time-series features were preprocessed to derive the representative values within an 8-hour time window. The values recorded closest to the specific time point and the recording frequencies over 8 hours prior to the time point were used as the representative values. In addition, the maximum, minimum, mean, and standard deviation values over 8 hours were calculated for numerical features (eg, the RASS, the GCS, and limb motor powers), and the recording frequencies for each category over 8 hours were considered for categorical features (eg, the CAM-ICU). We normalized the range of numerical features using a standardization method, which makes them have zero-mean and unit variance. We computed the parameters for normalization in the development sets and applied them to the full data sets.

Undersampling in the Data Sets

The number of UEs was scarce compared to planned extubation, resulting in an imbalance between the case and control numbers. To prevent overfitting of the control data, we undersampled the control 1 group using a simple random-sampling method and the control 2 group (ie, data from the planned extubation group) using a proportional stratified-sampling method. The days when the UE patients were on mechanical ventilation in the data sets were categorized into four groups. Control 2 data were sampled to thrice that of case data, while preserving the same proportion of days on mechanical ventilation for UE patients, as shown in Table 1. The sampled control data were independent, and the ratio of case to control 1 to control 2 in the data sets was approximately 1:1:3.

Table 1. Detailed information about unplanned extubation (UE) patients on mechanical ventilation that was used when undersampling the control 2 group.

Days on mechanical ventilation for UE patients[a]	Value (n=248), n (%)
1-2 days	83 (33.5)
2-3 days	53 (21.4)
3-5 days	55 (22.2)
>5 days	57 (23.0)

[a] Days when UE patients were on mechanical ventilation in the data sets were categorized into four groups.

Handling of Missing Data

We excluded 0.35% of the data where the RASS, the GCS, and limb motor powers were not recorded at least once in the whole time-series data sets. In terms of the features, the nearest value of the CAM-ICU was missing when there was no CAM-ICU record after ICU admission, where the missing rate was 33.46%. The missing data were assessed as *missing not at random* because the CAM-ICU was introduced to the hospital in which the study was conducted in late 2011 [18]. The CAM-ICU data were available after the method was introduced to the hospital, and there were many missing data at the beginning. We treated these data as a separate category altogether [19]. No missing data were estimated in the other features.

Feature Selection

Backward elimination, a stepwise approach, was used for feature selection. The random forest (RF) algorithm was applied to all the features, and the least important features, based on the measured predictor importance, were excluded [20]. Finally, a subset of features that optimized area under the receiver operating characteristic curve (AUROC) values was selected to develop the UE prediction models. AUROC scores that were based on varying numbers of features selected are shown in Multimedia Appendix 1. A total of 50 selected features as input of the models and their importance values are shown in Multimedia Appendix 2. The features and their importance values are plotted in Multimedia Appendix 3.

Modeling

Machine Learning Models

The following models were used to develop the UE prediction models: support vector machine (SVM), artificial neural network (ANN), logistic regression (LR), and RF [21-24].

Parameter Tuning

The parameters for SVM with the radial basis function kernel, LR, and RF models were tuned using grid search processes in the development sets, where the parameters with the best AUROC performance were selected. The hyperparameters for ANN, such as the number of layers and nodes in each layer, were tuned empirically. We used a five-layer network, with hidden layers having three to five times more neurons compared to the input features. For the activation function, a rectified linear unit was used in the hidden layer and a sigmoid function was used for the output layer [25]. To prevent the ANN from overfitting, we applied L2 regularization and dropout regularization [26,27]. The network was trained using mini-batch gradient descent and optimized using the cross-entropy method [28,29].

Validation

Initially, we conducted internal validation on the development sets to quantify optimism in the predictive performance and evaluate stability of the prediction model. Three repeated and stratified 5-fold cross-validation techniques were used to evaluate the internal validity of each model. In brief, the data set was randomly divided into five parts of roughly equal size, while preserving the ratio of cases and controls. When one part was used for validation, the remaining four parts were used for model training, where each prediction was summarized into the AUROC. This procedure, as mentioned above, was repeated three times.

Prior to validating the machine learning models based on the validation sets, thresholds for each model were determined. Three repeated and stratified 5-fold cross-validations were used in the development sets to identify the best threshold. The mean of 15 sensitivities and the mean of 15 specificities were calculated at thresholds from 0 to 1 with 0.005 units. The selected thresholds for each model had a mean sensitivity over 0.85, and the best threshold was identified to be the one with the highest mean specificity. Finally, the models were applied to the validation sets.

Statistical Analysis

Continuous variables were reported either as means and SDs for normal distribution data or as medians and IQRs for nonnormal distribution data. Categorical variables were reported as frequencies and percentages. We used the *t* test, the chi-square test, and the Wilcoxon rank-sum test to calculate the *P* values between the groups, where *P*<.05 was considered statistically significant.

The internal validation performance was evaluated through means and 95% CIs of the AUROCs. The performance of each model on the validation sets was evaluated with the AUROC, along with sensitivity, specificity, negative predictive value (NPV), positive predictive value (PPV), and the F1 score at the selected threshold.

For performance evaluation of the prediction model, we used a calibration curve, the Brier score [18,30], and the integrated calibration index (ICI) [31]. The potential clinical usefulness of the final model at the best threshold was assessed through a net-benefit approach using a decision curve [32]. This helps in determining if basing clinical decisions on a model is recommended considering the harm that it might cause, if any, in clinical practice. For statistical analyses and modeling, R, version 3.6.0 (The R Foundation) [33], and Python, version 3.6.6 (Python Software Foundation), were used [34]. The codes for developing and validating the models are available online [35].

Sample Size

The data sample for a diagnostic model should have an appropriate size [36]. Since there was no previous study that could directly be referred to, this study followed an often-used “rule of thumb,” where the sample size ensured at least 10 events per candidate predictor parameter [37,38]. The number of presumed events per candidate predictor in this study was 15, satisfying the rule.

Ethics Approval

The Institutional Review Board (IRB) of Samsung Medical Center approved this study (IRB file No. 2019-09-025).

Results

Study Population

A total of 6914 extubation cases that had occurred between January 1, 2010, and December 31, 2018, were included in the study. The flow diagram of the participant selection process is shown in Figure 2.

The basic characteristics of the included cases are listed in Table 2. During the study period, the occurrence of 248 UEs were reported. There were more males than females in the UE group. The UE group also had fewer surgical patients and a high proportion of patients with physical restraints. Both ICU mortality and hospital mortality were significantly higher in the UE group than in the planned extubation group. Further, the rate of reintubation within 24 hours was higher in the UE group. However, no differences were noted between groups regarding the length of mechanical ventilation.

Table 3 lists the characteristics of the development and validation sets. In the case group, where a UE event occurred, the recording frequency of the RASS over the last 8 hours, a RASS score over 2, eye and motor responses of the GCS, upper-limb motor power, lower-limb motor power, and the rate of physical restraint use were higher than in the control group for both the development and validation sets. The *missing rate* of CAM-ICU data in the validation sets was noticeably lower than in the development sets.

Figure 2. Flow diagram of the participant selection process. ICU: intensive care unit.

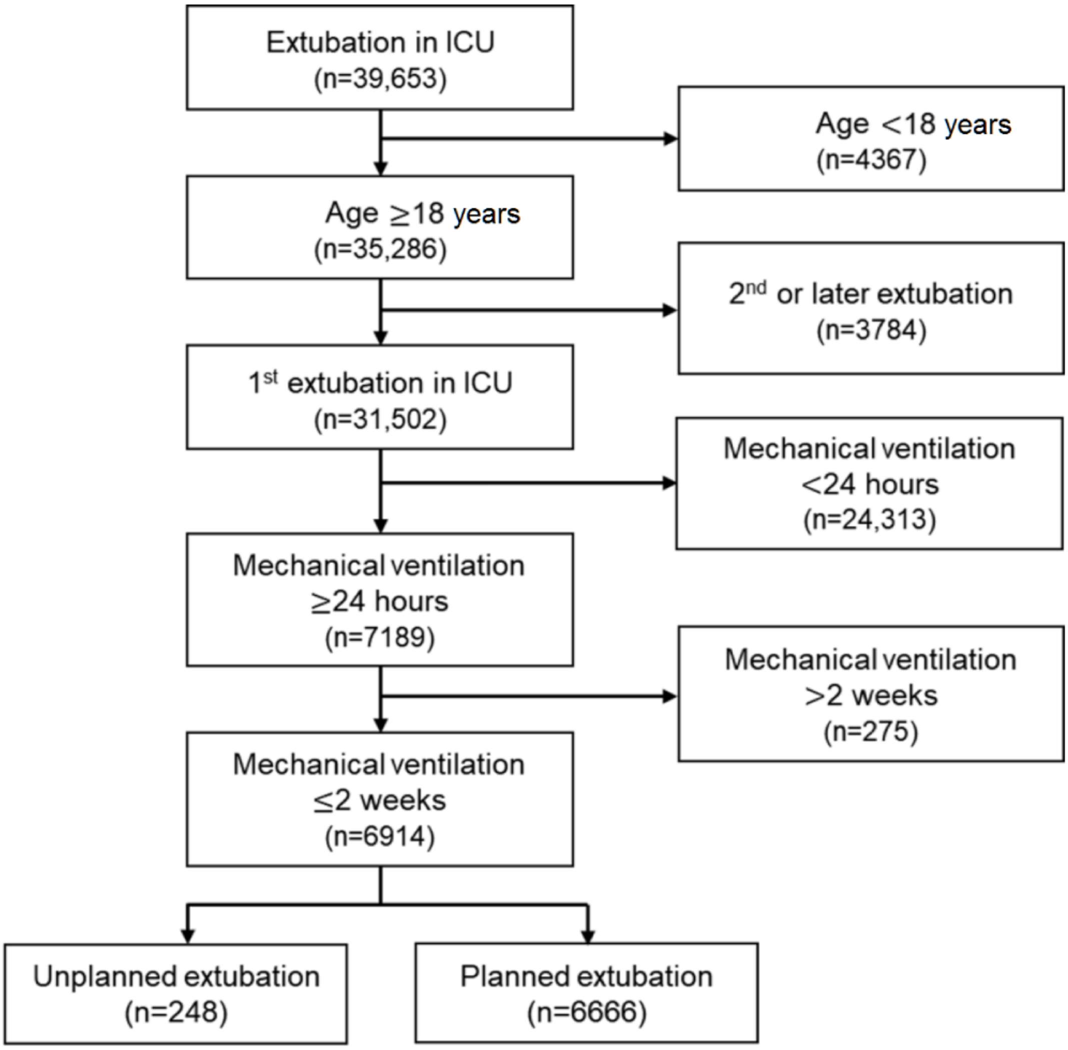

Extubation in ICU
(n=39,653)
Age <18 years
(n=4367)
Age ≥18 years
(n=35,286)
2nd or later extubation
(n=3784)
1st extubation in ICU
(n=31,502)
Mechanical ventilation
<24 hours
(n=24,313)
Mechanical ventilation
≥24 hours
(n=7189)
Mechanical ventilation
>2 weeks
(n=275)
Mechanical ventilation
≤2 weeks
(n=6914)
Unplanned extubation
(n=248)
Planned extubation
(n=6666)

Table 2. Basic characteristics and outcomes of the study population.

Characteristics and outcomes	Unplanned extubation (n=248)	Planned extubation (n=6666)	*P* value
Age (years), mean (SD)	62.2 (13.8)	62.1 (14.9)	.97
Sex, n (%)			<.001
Male	190 (76.6)	4319 (64.8)	
Female	58 (23.4)	2347 (35.2)	
Cardiopulmonary resuscitation, n (%)			.32
No	241 (97.2)	6377 (95.7)	
Yes	7 (2.8)	289 (4.3)	
Surgery, n (%)			<.001
No	184 (74.2)	3471 (52.1)	
Yes	64 (25.8)	3195 (47.9)	
Intubation location, n (%)			<.001
Emergency room	33 (13.3)	611 (9.2)	
Intensive care unit (ICU)	176 (71.0)	3997 (60.0)	
Operating room	17 (6.9)	1298 (19.5)	
Ward or others	22 (8.9)	760 (11.4)	
Reason for ICU admission, n (%)			<.001
Respiratory	138 (55.6)	2459 (36.9)	
Cardiovascular	41 (16.5)	909 (13.6)	
Perioperative	38 (15.3)	2345 (35.2)	
Others	31 (12.5)	953 (14.3)	
Use of physical restraint, n (%)[a]			<.001
No	96 (38.7)	4275 (64.1)	
Yes	152 (61.3)	2391 (35.9)	
Work shift, n (%)			<.001
Day (7 AM to 3 PM)	94 (37.9)	4121 (61.8)	
Evening (3 PM to 11 PM)	62 (25.0)	2123 (31.8)	
Night (11 PM to 7 AM)	92 (37.1)	422 (6.3)	
ICU mortality, n (%)			<.001
No	198 (79.8)	5847 (87.7)	
Yes	50 (20.2)	819 (12.3)	
In-hospital mortality, n (%)			<.001
No	150 (60.5)	4792 (71.9)	
Yes	98 (39.5)	1847 (28.1)	
Reintubation within 24 hours, n (%)			<.001
No	149 (60.1)	6128 (91.9)	
Yes	99 (39.9)	538 (8.1)	
Mechanical ventilation days, median (IQR)	2.7 (3.3)	2.9 (4.0)	.17
Hospital days, median (IQR)	27.5 (32.3)	25.0 (33.9)	.29

[a] Use of physical restraint indicates whether physical restraint was applied in a case when extubated.

Table 3. Characteristics of the development and validation sets.

Characteristics[a]	Development sets (n=1004)		Validation sets (n=191)	
	Case (n=209)	Control (n=795)	Case (n=39)	Control (n=152)
Age (years), mean (SD)	61.43 (13.86)	61.85 (14.39)	66.10 (13.13)	63.71 (14.97)
Sex, n (%)				
Male	159 (76.1)	522 (65.7)	31 (79.5)	100 (65.8)
Female	50 (23.9)	273 (34.3)	8 (20.5)	52 (34.2)
Surgery, n (%)				
No	52 (24.9)	294 (37.0)	5 (12.8)	30 (19.7)
Yes	157 (75.1)	501 (63.0)	34 (87.2)	122 (80.3)
Intubation location, n (%)				
Emergency room	26 (12.4)	61 (7.7)	7 (17.9)	24 (15.8)
Intensive care unit (ICU)	149 (71.3)	541 (68.1)	27 (69.2)	107 (70.4)
Operating room	15 (7.2)	94 (11.8)	2 (5.1)	13 (8.6)
Ward or others	19 (9.1)	99 (12.5)	3 (7.7)	8 (5.3)
Reason for ICU admission, n (%)				
Respiratory	36 (17.2)	237 (29.8)	2 (5.1)	24 (15.8)
Cardiovascular	30 (14.4)	108 (13.6)	11 (28.2)	29 (19.1)
Perioperative	36 (17.2)	237 (29.8)	2 (5.1)	24 (15.8)
Others	28 (13.4)	109 (13.7)	3 (7.7)	32 (21.1)
Recording frequency, mean (SD)				
Confusion Assessment Method for the Intensive Care Unit (CAM-ICU)	0.65 (0.63)	0.55 (0.53)	1.15 (0.43)	0.99 (0.45)
Richmond Agitation-Sedation Scale (RASS)	3.75 (5.93)	2.02 (2.84)	3.69 (3.64)	2.28 (2.72)
Glasgow Coma Scale (GCS)	3.38 (1.93)	3.52 (2.16)	2.59 (0.85)	2.91 (1.74)
Upper-limb motor power	3.01 (1.79)	3.18 (2.19)	2.54 (1.05)	2.79 (1.77)
Lower-limb motor power	3.01 (1.79)	3.17 (2.19)	2.51 (1.05)	2.79 (1.77)
Use of physical restraint	1.00 (1.07)	0.61 (0.84)	0.95 (0.69)	0.59 (0.67)
Nearest value of CAM-ICU, n (%)				
Negative	49 (23.4)	148 (18.6)	14 (35.9)	48 (31.6)
Positive	59 (28.2)	151 (19.0)	24 (61.5)	57 (37.5)
Unable to access	18 (8.6)	135 (17.0)	1 (2.6)	38 (25.0)
Missing	83 (39.7)	361 (45.4)	0 (0)	9 (5.9)
Nearest value of RASS, n (%)				
less than –2	20 (9.8)	213 (26.9)	2 (5.6)	42 (28.2)
–2 or –1	31 (15.1)	166 (20.9)	7 (19.4)	43 (28.9)
0	51 (24.9)	190 (24.0)	9 (25.0)	26 (17.4)
+1 or +2	60 (29.3)	163 (20.6)	4 (11.1)	25 (16.8)
more than +2	43 (21.0)	61 (7.7)	14 (38.9)	13 (8.7)
Nearest value of GCS, mean (SD)				
Eye response	3.38 (0.93)	2.93 (1.15)	3.54 (0.68)	3.04 (1.08)
Motor response	5.49 (1.19)	4.86 (1.71)	5.79 (0.52)	5.05 (1.60)
Nearest value of upper-limb motor power, mean (SD)				
Right	3.70 (1.31)	3.04 (1.62)	4.00 (0.76)	2.99 (1.59)
Left	3.72 (1.26)	3.05 (1.62)	4.10 (0.60)	3.00 (1.59)

Characteristics[a]	Development sets (n=1004)		Validation sets (n=191)	
	Case (n=209)	Control (n=795)	Case (n=39)	Control (n=152)
Nearest value of lower-limb motor power, mean (SD)				
Right	3.00 (1.48)	2.59 (1.62)	3.44 (1.05)	2.59 (1.54)
Left	3.03 (1.48)	2.62 (1.61)	3.44 (1.05)	2.62 (1.56)
Nearest value of use of physical restraint, n (%)				
No	84 (40.2)	460 (57.9)	13 (33.3)	79 (52.0)
Yes	125 (59.8)	335 (42.1)	26 (66.7)	73 (48.0)
Work shift, n (%)				
Day (7 AM to 3 PM)	78 (37.3)	296 (37.2)	16 (41.0)	61 (40.1)
Evening (3 PM to 11 PM)	49 (23.4)	242 (30.4)	13 (33.3)	38 (25.0)
Night (11 PM to 7 AM)	82 (39.2)	257 (32.3)	10 (25.6)	53 (34.9)

[a] For time-series features, the recording frequency over 8 hours prior to the time point and the nearest value to the time point were derived.

Model Development and Assessment

A total of 50 features, selected through a recursive feature-elimination technique among the 66 candidates, reflected demographic characteristics and patterns of change in the time-series data. The features, their importance scores, and their variable types are listed in Multimedia Appendix 2. The list of the selected features with their corresponding importance scores are plotted in Multimedia Appendix 3.

We developed machine learning–based prediction algorithms using RF, LR, ANN, and SVM. The average AUROCs and 95% CIs for internal validation in the development sets were 0.732 (95% CI 0.705-0.759) for RF, 0.703 (95% CI 0.676-0.730) for LR, 0.670 (95% CI 0.637-0.702) for ANN, and 0.689 (95% CI 0.668-0.710) for SVM.

For each model, the highest value of specificity among the sensitivities over 0.85 was selected as the cutoff point of the threshold. In terms of the machine learning models, the best model was RF, with the highest performance values at the selected threshold, where AUROC was 0.787 and sensitivity, specificity, NPV, PPV, F1 score, and ICI were 0.949, 0.388, 0.967, 0.285, 0.438, and 0.048, respectively. The performance values of the prediction models are listed in Table 4. The models' AUROCs are shown in Figure 3.

The performance of the best model was evaluated using the Brier score, the ICI, and decision curve analysis. The calibration, agreement between observed outcomes and predicted risk probabilities, was assessed with the slope of the calibration curve and the Brier score. The RF model was well-calibrated, and the Brier score and ICI were 0.129 and 0.048, respectively. The calibration curve of the best model is shown in Figure 4. The decision curve compared the net benefit of the best model and alternative approaches for clinical decision making. The decision curve showed superior net benefit when the best model was used compared to the alternative approaches of "predicting all as a UE" or "predicting none as a UE" over a threshold probability range of 6% to 78%. Our selected threshold was 14%, and it showed potentially superior clinical utility. The decision curve of the best model is presented in Figure 5.

Table 4. Comparison of performance values of the prediction models.

Model	AUROC[a]	Sensitivity	Specificity	NPV[b]	PPV[c]	F1 score	ICI[d]
Random forest	0.787	0.949	0.388	0.967	0.285	0.438	0.048
Linear regression	0.762	0.949	0.303	0.958	0.259	0.407	0.025
Artificial neural network	0.763	0.949	0.230	0.946	0.240	0.383	0.077
Support vector machine	0.740	0.897	0.283	0.915	0.243	0.383	0.050

[a] AUROC: area under the receiver operating characteristic curve.

[b] NPV: negative predictive value.

[c] PPV: positive predictive value.

[d] ICI: integrated calibration index.

Figure 3. Receiver operating characteristic curves for all of the unplanned extubation prediction models. ANN: artificial neural network; AUROC: area under the receiver operating characteristic curve; LR: linear regression; RF: random forest; SVM: support vector machine.

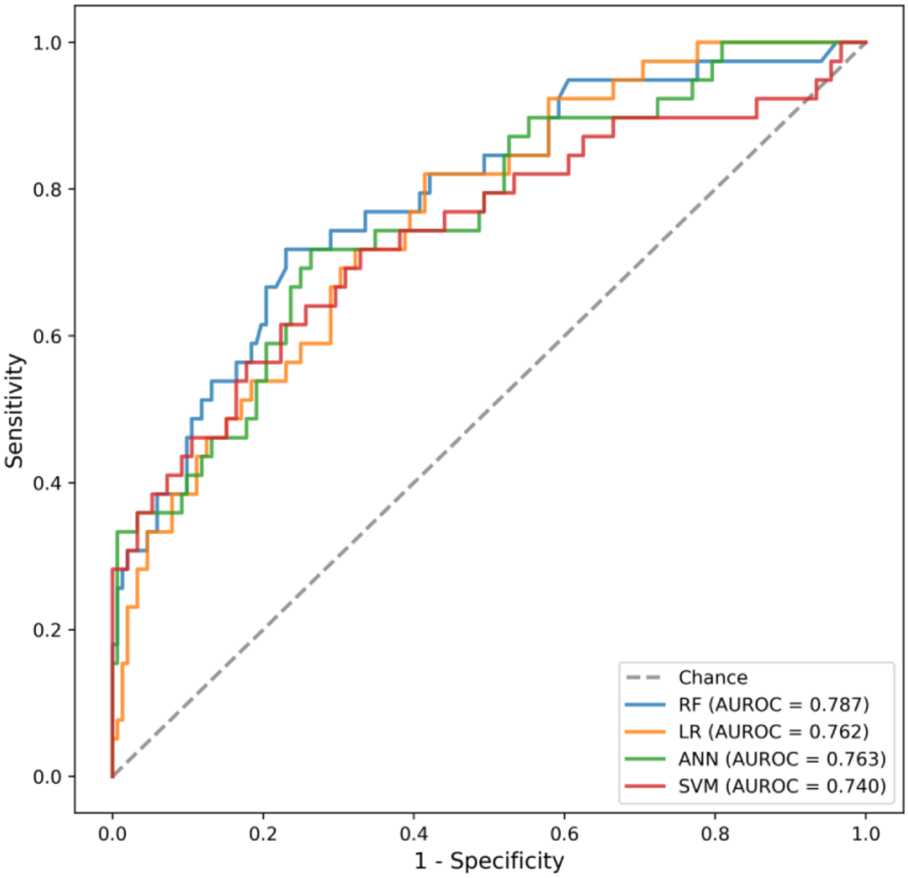


Figure 4. The calibration curve and the integrated calibration index (ICI) of the best model.

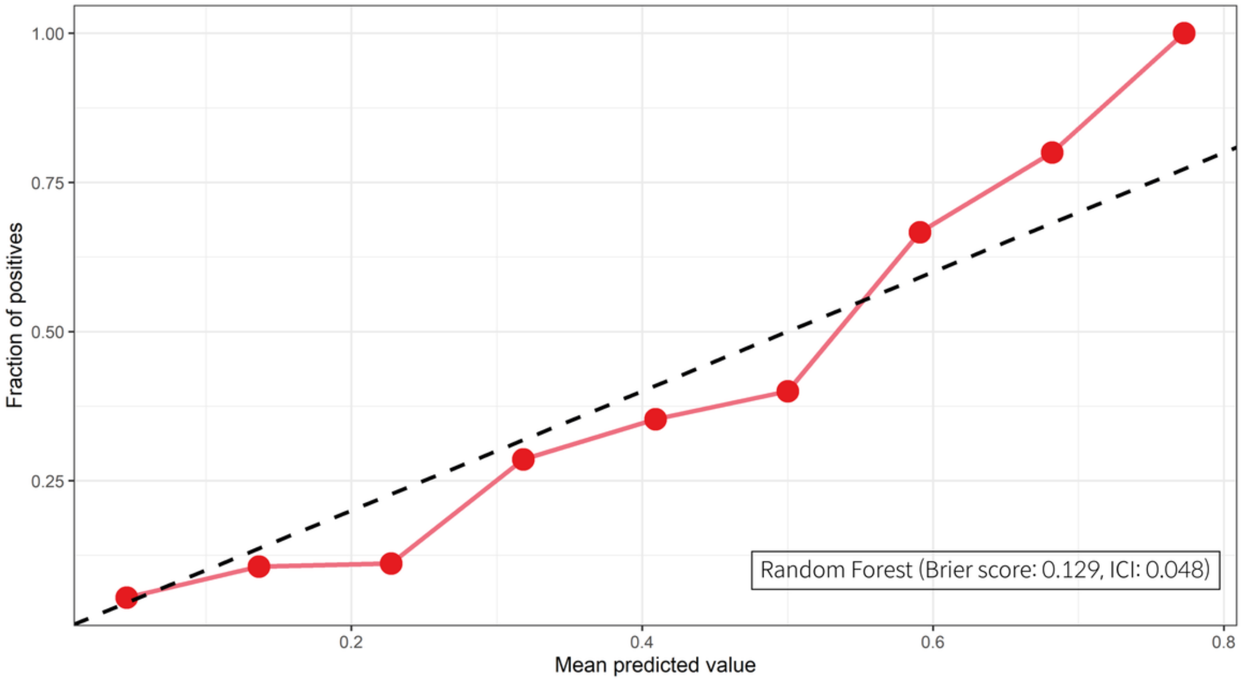
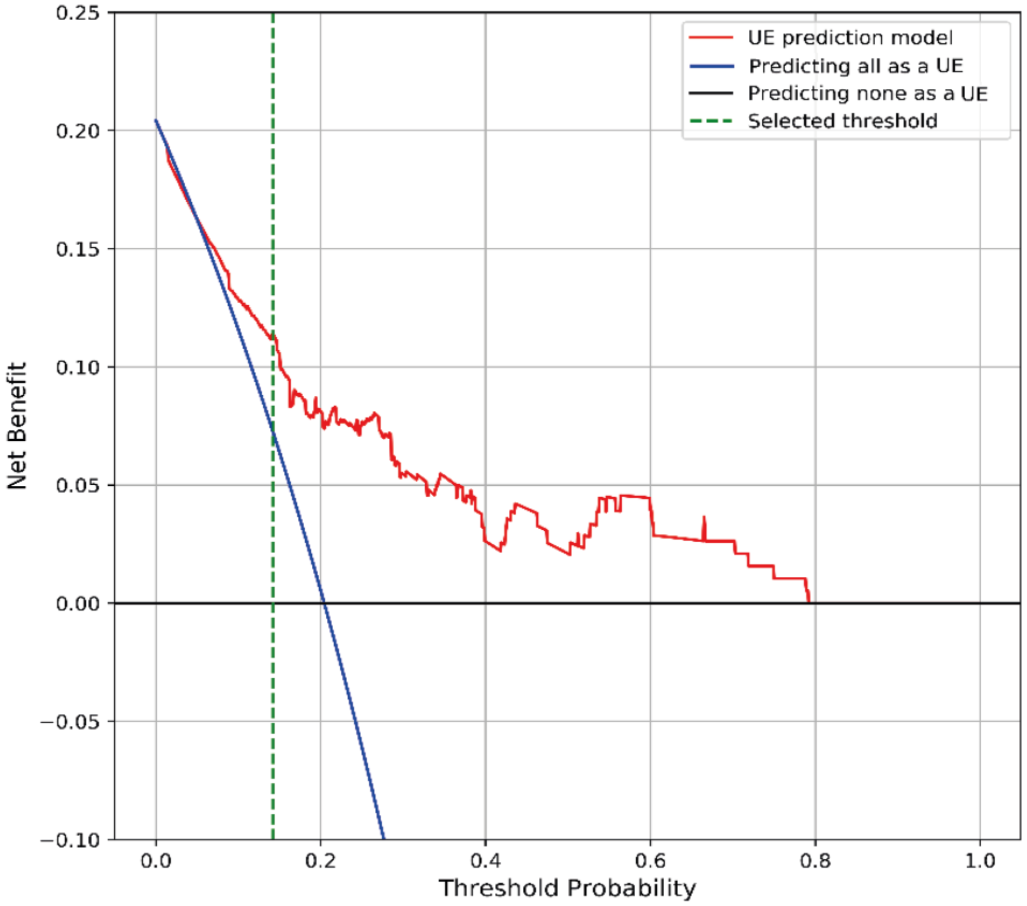


Figure 5. The decision curve of the best model. UE: unplanned extubation.



Discussion

Principal Findings

For patient safety, prevention and early detection of clinical error is an essential component of high-quality care [1]. The proposed prediction model is expected to screen and monitor ICU patients effectively when applied to the clinical setting. To the best of our knowledge, this is the first machine learning–based prediction model for UE incidents, and it is an algorithm that predicts the UE within 1 hour, allowing clinical staff to take appropriate action to prevent UE. In the previous study, a simple LR-based statistical model was presented where the data were not divided into training and test sets [39].

The limitation of the machine learning prediction model is related to its ability to exhibit good performance in a real clinical setting. Our study assessed the performance of the UE prediction model; the best model demonstrated good calibration and net benefit over a wide range of threshold probabilities. This prediction model shows potentially superior clinical utility based on decision curve analysis [40].

Comparison With Prior Work

Existing UE risk assessment tools and applications will have a limited impact if they include additional work for the nurses, such as requiring additional assessments or documentation tasks. An EHR-based prediction algorithm can automatically calculate the risk for clinical staff without any additional workload. Alarm fatigue in the ICU is another major concern that disrupts the workflow of the clinician and can significantly impact patient safety [41]. The UE prediction model is intended to be used as a screening tool for predicting potential UE events, otherwise the false alarm rate would be high due to the low specificity and PPV [42]. Therefore, clinician stakeholders would need to be engaged in identifying ways to ensure that the alert is integrated into the clinical workflow in a way that is actionable. Clinicians should also be involved in setting appropriate threshold values based on their practice, workflow, and purpose for adopting the algorithm [43].

In previous studies, agitation was the most important factor among patient-associated risk factors for UE incidence. The incidence rate of UE varies according to the patient's level of consciousness, recording frequency, and age; in addition, physical restraints were significant risk factors for UE (Multimedia Appendix 2). Recording frequency is presented as an important feature, and frequent recording of the patient's condition in clinical practice provides an interpretation that improves predictions.

Further, this study revealed that the use of physical restraints was higher in the UE group. Though physical restraints are frequently used in ICUs to prevent UE [44,45], it can increase the risk of UE [46]. A factor that can be attributed to this ironic result is the use of restraints evoking delirium, which is related to self-extubation [47]. However, the physical restraints may have been warranted as a safety measure, but insufficiently applied and, therefore, unable to prevent UE.

Limitations

This study was retrospective and carried out in a single center. To improve the model’s performance and for precise comparison among machine learning–based models, comparatively large clinical data sets and multicenter validation are required. All developed models seemed to have similar performances, assuming that small evaluation data sets caused this. Further, prospective studies are required to verify the algorithm’s performance.

There are limitations in terms of the number of small data sets and random sampling for the control 2 group, resulting in a biased sample. Although UE is a significant complication in the ICU, its incident rate was reported to be low in the previous studies. Thus, it is complicated to obtain large amounts of data on events related to patient safety accidents. Obtaining ample data is a crucial concern in machine learning. Validating a prediction model requires a minimum of 100 events and 100 nonevents; however, our validation data set did not include 100 events. Instead, our study had 15 events per candidate predictor in the development data set and satisfied the well-used “rule of thumb.” Nevertheless, machine learning is possible with the use of small data sets [48-50]. We conducted a stratified undersampling method to avoid overfitting, and data were sampled randomly. This method can potentially discard important information or results in a biased sample.

In this study, we included short-term mortality (ie, ICU mortality) and in-hospital mortality. We could not follow up on deaths of patients after discharge. Further, we have not considered long-term survival and correlation between comorbidity and duration of mechanical ventilation.

Future Research

The models were developed retrospectively and carried out in a single center; therefore, future prospective evaluation and validation using other data sets are required.

Conclusions

We developed a machine learning prediction model for UE patients. The best AUROC was 0.787, and the sensitivity was 0.949 at the selected threshold for the best model. The best model was well-calibrated, and the Brier score and ICI were 0.129 and 0.048, respectively. The proposed prediction model uses widely available variables to limit the additional workload on the clinician. Further, this evaluation suggests that the model holds potential for clinical usefulness.

Acknowledgments

This research was supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute, funded by the Ministry of Health and Welfare, Republic of Korea (grant HI19C0275).

Authors' Contributions

SH designed the study, extracted and analyzed the data, and wrote the paper as the first author. JYM designed the study, analyzed the data, and wrote the paper as the co–first author. KK contributed to the analysis of the results in a statistical aspect. CRC assisted in the support of clinical knowledge and reviewed the paper. PCD contributed to the analysis of the results and reviewed the paper. WCC was in charge of the overall direction of the study as the corresponding author. All authors gave final approval of the paper for submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Area under the receiver operator characteristic curve (AUROC) scores based on varying numbers of features selected. A stepwise backward elimination technique (recursive feature elimination [RFE])—based on feature importance derived from the random forest algorithm with 500 trees and three repeated and stratified 5-fold cross-validation techniques—was used to select the optimal subset of features. The feature subset scores were based on the mean of the AUROCs from cross-validation. The RFE with cross-validation (RFECV) function in the scikit-learn package, version 0.22.1, was used for feature selection.

[PNG File , 76 KB - jmir_v23i8e23508_app1.png]

Multimedia Appendix 2

The 50 selected features as input of the models.

[DOCX File , 17 KB - jmir_v23i8e23508_app2.docx]

Multimedia Appendix 3

Importance of features included in the unplanned extubation prediction models following application of the random forest algorithm.

[PNG File , 518 KB - jmir_v23i8e23508_app3.png]

References

1. Valentin A, Capuzzo M, Guidet B, Moreno RP, Dolanski L, Bauer P, Research Group on Quality Improvement of European Society of Intensive Care Medicine, Sentinel Events Evaluation Study Investigators. Patient safety in intensive care: Results from the multinational Sentinel Events Evaluation (SEE) study. Intensive Care Med 2006 Oct;32(10):1591-1598. [doi: 10.1007/s00134-006-0290-7] [Medline: 16874492]
2. Chacko J, Raju HR, Singh MK, Mishra RC. Critical incidents in a multidisciplinary intensive care unit. Anaesth Intensive Care 2007 Jun;35(3):382-386. [doi: 10.1177/0310057X0703500311] [Medline: 17591133]
3. Epstein SK, Nevins ML, Chung J. Effect of unplanned extubation on outcome of mechanical ventilation. Am J Respir Crit Care Med 2000 Jun;161(6):1912-1916. [doi: 10.1164/ajrccm.161.6.9908068] [Medline: 10852766]
4. Kapadia FN, Bajan KB, Raje KV. Airway accidents in intubated intensive care unit patients: An epidemiological study. Crit Care Med 2000 Mar;28(3):659-664. [doi: 10.1097/00003246-200003000-00010] [Medline: 10752811]
5. Balon JA. Common factors of spontaneous self-extubation in a critical care setting. Int J Trauma Nurs 2001;7(3):93-99. [doi: 10.1067/mtn.2001.117769] [Medline: 11477388]
6. Lucchini A, Bambi S, Galazzi A, Elli S, Negrini C, Vaccino S, et al. Unplanned extubations in general intensive care unit: A nine-year retrospective analysis. Acta Biomed 2018 Dec 07;89(7-S):25-31 [FREE Full text] [doi: 10.23750/abm.v89i7-S.7815] [Medline: 30539936]
7. Atkins PM, Mion LC, Mendelson W, Palmer RM, Slomka J, Franko T. Characteristics and outcomes of patients who self-extubate from ventilatory support: A case-control study. Chest 1997 Nov 05;112(5):1317-1323. [doi: 10.1378/chest.112.5.1317] [Medline: 9367475]
8. Boulain T. Unplanned extubations in the adult intensive care unit: A prospective multicenter study. Association des Réanimateurs du Centre-Ouest. Am J Respir Crit Care Med 1998 Apr;157(4 Pt 1):1131-1137. [doi: 10.1164/ajrccm.157.4.9702083] [Medline: 9563730]
9. Lin CC, Wu HL, Chen SY. A new device to prevent unplanned endotracheal self-extubation in mechanically ventilated patients. In: Proceedings of the European Respiratory Society (ERS) International Congress. 2018 Nov 19 Presented at: European Respiratory Society (ERS) International Congress; September 15-19, 2018; Paris, France. [doi: 10.1183/13993003.congress-2018.pa2339]
10. Ferraz P, Barros M, Miyoshi M, Davidson J, Guinsburg R. Bundle to reduce unplanned extubation in a neonatal intensive care unit. J Matern Fetal Neonatal Med 2020 Sep;33(18):3077-3085. [doi: 10.1080/14767058.2019.1568981] [Medline: 30632822]
11. 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]
12. Ford DW, Goodwin AJ, Simpson AN, Johnson E, Nadig N, Simpson KN. A severe sepsis mortality prediction model and score for use with administrative data. Crit Care Med 2016 Feb;44(2):319-327 [FREE Full text] [doi: 10.1097/CCM.0000000000001392] [Medline: 26496452]
13. Golas SB, Shibahara T, Agboola S, Otaki H, Sato J, Nakae T, et al. A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: A retrospective analysis of electronic medical records data. BMC Med Inform Decis Mak 2018 Jun 22;18(1):44 [FREE Full text] [doi: 10.1186/s12911-018-0620-z] [Medline: 29929496]
14. Ong MEH, Lee Ng CH, Goh K, Liu N, Koh ZX, Shahidah N, et al. Prediction of cardiac arrest in critically ill patients presenting to the emergency department using a machine learning score incorporating heart rate variability compared with the modified early warning score. Crit Care 2012 Jun 21;16(3):R108 [FREE Full text] [doi: 10.1186/cc11396] [Medline: 22715923]
15. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD). Circulation 2015 Jan 13;131(2):211-219. [doi: 10.1161/circulationaha.114.014508]
16. Ely EW, Margolin R, Francis J, May L, Truman B, Dittus R, et al. Evaluation of delirium in critically ill patients: Validation of the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU). Crit Care Med 2001 Jul;29(7):1370-1379. [doi: 10.1097/00003246-200107000-00012] [Medline: 11445689]
17. Sessler CN, Gosnell MS, Grap MJ, Brophy GM, O'Neal PV, Keane KA, et al. The Richmond Agitation-Sedation Scale: Validity and reliability in adult intensive care unit patients. Am J Respir Crit Care Med 2002 Nov 15;166(10):1338-1344. [doi: 10.1164/rccm.2107138] [Medline: 12421743]
18. Rubin DB. Inference and missing data. Biometrika 1976;63(3):581-592. [doi: 10.1093/biomet/63.3.581]
19. Greenland S, Finkle WD. A critical look at methods for handling missing covariates in epidemiologic regression analyses. Am J Epidemiol 1995 Dec 15;142(12):1255-1264. [doi: 10.1093/oxfordjournals.aje.a117592] [Medline: 7503045]
20. Saeys Y, Inza I, Larrañaga P. A review of feature selection techniques in bioinformatics. Bioinformatics 2007 Oct 01;23(19):2507-2517. [doi: 10.1093/bioinformatics/btm344] [Medline: 17720704]
21. Suykens JAK, Vandewalle J. Least squares support vector machine classifiers. Neural Process Lett 1999 Jun;9(3):293-300. [doi: 10.1023/A:1018628609742]
22. Rosenblatt F. The perceptron: A probabilistic model for information storage and organization in the brain. Psychol Rev 1958 Nov;65(6):386-408. [doi: 10.1037/h0042519] [Medline: 13602029]

23. Hosmer Jr DW, Lemeshow S, Sturdivant RX. Applied Logistic Regression. 3rd edition. Hoboken, NJ: John Wiley & Sons; 2013.
24. Breiman L. Random forests. Mach Learn 2001;45:5-32. [doi: 10.1023/A:1010933404324]
25. Nair V, Hinton G. Rectified linear units improve restricted Boltzmann machines. In: Proceedings of the 27th International Conference on Machine Learning. 2010 Presented at: 27th International Conference on Machine Learning; June 21-24, 2010; Haifa, Israel URL: https://www.cs.toronto.edu/~fritz/absps/reluICML.pdf
26. Ng AY. Feature selection, L1 vs L2 regularization, and rotational invariance. In: Proceedings of the 21st International Conference on Machine Learning. 2004 Presented at: 21st International Conference on Machine Learning; July 4-8, 2004; Banff, AB. [doi: 10.1145/1015330.1015435]
27. Hinton G, Srivastava N, Krizhevsky A, Sutskever I, Salakhutdinov R. Improving neural networks by preventing co-adaptation of feature detectors. ArXiv. Preprint posted online on July 3, 2012 [FREE Full text]
28. Khirirat S, Feyzmahdavian H, Johansson M. Mini-batch gradient descent: Faster convergence under data sparsity. In: Proceedings of the 2017 IEEE 56th Annual Conference on Decision and Control.: IEEE; 2017 Presented at: 2017 IEEE 56th Annual Conference on Decision and Control; December 12-15, 2017; Melbourne, Australia p. 2880-2887. [doi: 10.1109/cdc.2017.8264077]
29. de Boer PT, Kroese DP, Mannor S, Rubinstein RY. A tutorial on the cross-entropy method. Ann Oper Res 2005 Feb;134(1):19-67. [doi: 10.1007/s10479-005-5724-z]
30. Gerds TA, Cai T, Schumacher M. The performance of risk prediction models. Biom J 2008 Aug;50(4):457-479. [doi: 10.1002/bimj.200810443] [Medline: 18663757]
31. Austin PC, Steyerberg EW. The Integrated Calibration Index (ICI) and related metrics for quantifying the calibration of logistic regression models. Stat Med 2019 Sep 20;38(21):4051-4065 [FREE Full text] [doi: 10.1002/sim.8281] [Medline: 31270850]
32. Vickers AJ, Elkin EB. Decision curve analysis: A novel method for evaluating prediction models. Med Decis Making 2006;26(6):565-574 [FREE Full text] [doi: 10.1177/0272989X06295361] [Medline: 17099194]
33. R-3.6.0 for Windows. The Comprehensive R Archive Network. Vienna, Austria: The R Foundation; 2019. URL: https://cran.r-project.org/bin/windows/base/old/3.6.0/ [accessed 2021-08-05]
34. van Rossum G, Drake FL. Python 3 Reference Manual. Scotts Valley, CA: CreateSpace; 2009.
35. UE prediction model. GitHub. 2020. URL: https://github.com/wnet500/UE_prediction_model [accessed 2021-08-05]
36. Riley RD, Snell KI, Ensor J, Burke DL, Harrell FE, Moons KG, et al. Minimum sample size for developing a multivariable prediction model: PART II - Binary and time-to-event outcomes. Stat Med 2019 Mar 30;38(7):1276-1296 [FREE Full text] [doi: 10.1002/sim.7992] [Medline: 30357870]
37. Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of events per independent variable in proportional hazards regression analysis II. Accuracy and precision of regression estimates. J Clin Epidemiol 1995 Dec;48(12):1503-1510. [doi: 10.1016/0895-4356(95)00048-8]
38. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol 1996 Dec;49(12):1373-1379. [doi: 10.1016/s0895-4356(96)00236-3]
39. Lee JY, Park H, Chung E. Use of electronic critical care flow sheet data to predict unplanned extubation in ICUs. Int J Med Inform 2018 Sep;117:6-12. [doi: 10.1016/j.ijmedinf.2018.05.011] [Medline: 30032966]
40. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. BMJ 2016 Jan 25;352:i6 [FREE Full text] [doi: 10.1136/bmj.i6] [Medline: 26810254]
41. Cvach M. Monitor alarm fatigue: An integrative review. Biomed Instrum Technol 2012;46(4):268-277. [doi: 10.2345/0899-8205-46.4.268] [Medline: 22839984]
42. van der Sijs H, Aarts J, Vulto A, Berg M. Overriding of drug safety alerts in computerized physician order entry. J Am Med Inform Assoc 2006;13(2):138-147 [FREE Full text] [doi: 10.1197/jamia.M1809] [Medline: 16357358]
43. Reilly BM, Evans AT. Translating clinical research into clinical practice: Impact of using prediction rules to make decisions. Ann Intern Med 2006 Feb 07;144(3):201-209. [doi: 10.7326/0003-4819-144-3-200602070-00009] [Medline: 16461965]
44. Tanios MA, Epstein SK, Livelo J, Teres D. Can we identify patients at high risk for unplanned extubation? A large-scale multidisciplinary survey. Respir Care 2010 May;55(5):561-568 [FREE Full text] [Medline: 20420726]
45. Benbenbishty J, Adam S, Endacott R. Physical restraint use in intensive care units across Europe: The PRICE study. Intensive Crit Care Nurs 2010 Oct;26(5):241-245. [doi: 10.1016/j.iccn.2010.08.003] [Medline: 20837320]
46. Chang L, Wang KK, Chao Y. Influence of physical restraint on unplanned extubation of adult intensive care patients: A case-control study. Am J Crit Care 2008 Sep;17(5):408-415; quiz 416. [Medline: 18775996]
47. Van Rompaey B, Elseviers MM, Schuurmans MJ, Shortridge-Baggett LM, Truijen S, Bossaert L. Risk factors for delirium in intensive care patients: A prospective cohort study. Crit Care 2009;13(3):R77 [FREE Full text] [doi: 10.1186/cc7892] [Medline: 19457226]
48. Shaikhina T, Lowe D, Daga S, Briggs D, Higgins R, Khovanova N. Decision tree and random forest models for outcome prediction in antibody incompatible kidney transplantation. Biomed Signal Process Control 2019 Jul;52:456-462. [doi: 10.1016/j.bspc.2017.01.012]

49. Segar MW, Vaduganathan M, Patel KV, McGuire DK, Butler J, Fonarow GC, et al. Machine learning to predict the risk of incident heart failure hospitalization among patients with diabetes: The WATCH-DM risk score. Diabetes Care 2019 Dec;42(12):2298-2306 [FREE Full text] [doi: 10.2337/dc19-0587] [Medline: 31519694]
50. Goto S, Kimura M, Katsumata Y, Goto S, Kamatani T, Ichihara G, et al. Artificial intelligence to predict needs for urgent revascularization from 12-leads electrocardiography in emergency patients. PLoS One 2019;14(1):e0210103 [FREE Full text] [doi: 10.1371/journal.pone.0210103] [Medline: 30625197]

Abbreviations

ANN: artificial neural network
AUROC: area under the receiver operating characteristic curve
CAM-ICU: Confusion Assessment Method for the Intensive Care Unit
EHR: electronic health record
GCS: Glasgow Coma Scale
ICI: integrated calibration index
ICU: intensive care unit
IRB: Institutional Review Board
LR: logistic regression
NPV: negative predictive value
PPV: positive predictive value
RASS: Richmond Agitation-Sedation Scale
RF: random forest
SVM: support vector machine
TRIPOD: Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis
UE: unplanned extubation

Edited by R Kukafka; submitted 16.08.20; peer-reviewed by R Bajpai, A Sharafoddini; comments to author 03.09.20; revised version received 19.10.20; accepted 13.07.21; published 11.08.21.

Please cite as:
Hur S, Min JY, Yoo J, Kim K, Chung CR, Dykes PC, Cha WC
Development and Validation of Unplanned Extubation Prediction Models Using Intensive Care Unit Data: Retrospective, Comparative, Machine Learning Study
J Med Internet Res 2021;23(8):e23508
URL: https://www.jmir.org/2021/8/e23508
doi:10.2196/23508
PMID:34382940

©Sujeong Hur, Ji Young Min, Junsang Yoo, Kyunga Kim, Chi Ryang Chung, Patricia C Dykes, Won Chul Cha. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 11.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

A Modified Public Health Automated Case Event Reporting Platform for Enhancing Electronic Laboratory Reports With Clinical Data: Design and Implementation Study

Ninad Mishra[1*], MS, MD; Jon Duke[2*], MS, MD; Saugat Karki[1*], MS, MD; Myung Choi[2*], PhD; Michael Riley[2*], BS; Andrey V Ilatovskiy[3*], PhD; Marla Gorges[2*], BS; Leslie Lenert[3*], MS, MD

[1]Division of STD Prevention, National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention, Centers for Disease Control and Prevention, Atlanta, GA, United States

[2]Center for Health Analytics and Informatics, Georgia Tech Research Institute, Atlanta, GA, United States

[3]Biomedical Informatics Center, Medical University of South Carolina, Charleston, SC, United States

[*]all authors contributed equally

Corresponding Author:
Saugat Karki, MS, MD
Division of STD Prevention
National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention
Centers for Disease Control and Prevention
US 12-2, Rm 2031
1600 Clifton Road
Atlanta, GA, 30329
United States
Phone: 1 4047187483
Email: oql7@cdc.gov

Abstract

Background: Public health reporting is the cornerstone of public health practices that inform prevention and control strategies. There is a need to leverage advances made in the past to implement an architecture that facilitates the timely and complete public health reporting of relevant case-related information that has previously not easily been available to the public health community. Electronic laboratory reporting (ELR) is a reliable method for reporting cases to public health authorities but contains very limited data. In an earlier pilot study, we designed the Public Health Automated Case Event Reporting (PACER) platform, which leverages existing ELR infrastructure as the trigger for creating an electronic case report. PACER is a FHIR (Fast Health Interoperability Resources)-based system that queries the electronic health record from where the laboratory test was requested to extract expanded additional information about a case.

Objective: This study aims to analyze the pilot implementation of a modified PACER system for electronic case reporting and describe how this FHIR-based, open-source, and interoperable system allows health systems to conduct public health reporting while maintaining the appropriate governance of the clinical data.

Methods: ELR to a simulated public health department was used as the trigger for a FHIR-based query. Predetermined queries were translated into Clinical Quality Language logics. Within the PACER environment, these Clinical Quality Language logical statements were managed and evaluated against the providers' FHIR servers. These predetermined logics were filtered, and only data relevant to that episode of the condition were extracted and sent to simulated public health agencies as an electronic case report. Design and testing were conducted at the Georgia Tech Research Institute, and the pilot was deployed at the Medical University of South Carolina. We evaluated this architecture by examining the completeness of additional information in the electronic case report, such as patient demographics, medications, symptoms, and diagnoses. This additional information is crucial for understanding disease epidemiology, but existing electronic case reporting and ELR architectures do not report them. Therefore, we used the completeness of these data fields as the metrics for enriching electronic case reports.

Results: During the 8-week study period, we identified 117 positive test results for chlamydia. PACER successfully created an electronic case report for all 117 patients. PACER extracted demographics, medications, symptoms, and diagnoses from 99.1% (116/117), 72.6% (85/117), 70.9% (83/117), and 65% (76/117) of the cases, respectively.

Conclusions: PACER deployed in conjunction with electronic laboratory reports can enhance public health case reporting with additional relevant data. The architecture is modular in design, thereby allowing it to be used for any reportable condition, including evolving outbreaks. PACER allows for the creation of an enhanced and more complete case report that contains relevant case information that helps us to better understand the epidemiology of a disease.

(J Med Internet Res 2021;23(8):e26388) doi:10.2196/26388

KEYWORDS

public health surveillance; sexually transmitted diseases; gonorrhea; chlamydia; electronic case reporting; electronic laboratory reporting; health information interoperability; fast healthcare interoperability resources; electronic health records; EHR

Introduction

Background

Public health surveillance—the ongoing and systematic collection, analysis, and interpretation of health-related data essential to planning, implementation, and evaluation of public health practice [1]—has been the cornerstone of public health practice. These data are important for understanding the burden, trends, pattern, and general epidemiology of diseases. These requirements for reporting diseases are mandated by state laws or regulations. In 1990, the Council for State and Territorial Epidemiologists (CSTE) and the US Centers for Disease Control and Prevention collaborated to develop uniform criteria to author case definitions and to facilitate reporting of nationally notifiable diseases [2]. To qualify a case as reportable, it must meet the case definition for the condition [3]. These case definition criteria are based on laboratory information, clinical observations, or both. Previously, notifiable diseases were manually reported to public health agencies. Advancements in data and messaging standards have led to the modernization of this process by electronically reporting cases from either electronic health records (EHRs) via electronic case reporting (eCR) or laboratories via electronic laboratory reporting (ELR) [4]. Health level 7 (HL7)–based messages for ELR and eCR have been developed as tools to facilitate case reporting, alleviate the burden of reporting from clinical providers, and establish communication between often disparate EHR systems [5]. Electronic laboratory reports are created on the HL7 version 2 specification [6]. ELR has been increasing in adoption and is a reliable source of reporting cases to public health authorities [7] but contains little information about a case other than the test result. This information, such as demographics, diagnoses, and treatment, is important to public health to understand disease epidemiology and typically exists within the EHR. Similarly, existing eCR architectures are created on the HL7 Consolidated Clinical Document Architecture (C-CDA) [8]. Although eCR is triggered from the EHR, treatment information is not always available at the time the electronic case report is created, which typically occurs earlier on in the continuum of care when the triggering criteria have been met for an automated case report to be sent to public health authorities. Demographic information is important to understand the epidemiology, which, in turn, informs the policies and strategies for the prevention and control of diseases. Treatment information is particularly important at this point in time because of the worrisome trend of antimicrobial resistance exhibited in gonorrhea cases. To address this emerging threat of antimicrobial-resistant gonorrhea and to ensure that patients receive the highest quality of care, monitoring of treatment practices is a critical public health priority [9].

Improvements in interoperability and automation are crucial for public health surveillance [10]. Health care data standards have been developed to reduce ambiguity and facilitate interoperability among health care providers, public health, and other stakeholders [11]. Advances have been made by messaging standards developed by global leaders, such as the HL7. Furthermore, the adoption of FHIR (Fast Health Interoperability Resources), an HL7 standard, has moved semantic interoperability forward by leaps over the last several years [12]. FHIR are easy to implement and leverage the latest web data exchange standards while using existing data standards such as Logical Observation Identifiers Names and Codes (LOINC) [13]; Systematized Nomenclature of Medicine—Clinical Terms [14]; and International Classification of Diseases, Tenth Revision, Clinical Modification [15]. FHIR *Resources* are building blocks that contain a common way to define and represent concepts, sometimes from external terminologies or ontologies (Systematized Nomenclature of Medicine—Clinical Terms, RxNorm, LOINC, etc), a common set of metadata, and a human-readable component [16]. In addition, Clinical Quality Language (CQL), an HL7 authoring standard, has been developed to harmonize standards used for authoring clinical decision support and clinical quality measurement artifacts [17]. CQL is intended to be human-readable and enables queries into EHRs via a set of curated logics. In an earlier pilot, we successfully demonstrated a FHIR-based case reporting at the Indiana Health Information Exchange, which automatically created electronic case reports in 84.6% of the cases [18]. This architecture of eCR extracted all clinical encounters associated with the patient, resulting in a large amount of redundant data that were not associated with that particular episode of infection. It was evident that we had to not only filter information related to the particular condition but also to allow health systems to have control over what type of data are reported to public health authorities.

Objectives

In this study, we evaluated the pilot implementation of a modified FHIR-based approach for eCR that leverages electronic laboratory reports as triggers to capture relevant expanded case data from EHRs. We describe how this interoperable approach, which is based on open-source and widely adopted health care data standards, can bridge public health and health care data systems by providing relevant, timely, and substantially more information on reportable cases. This implemented architecture, which is known as Public Health Automated Case Event

Reporting (PACER), leverages not only FHIR protocols but also the HL7 CQL for interoperable query design. This modified architecture is specifically designed to ensure that health systems have complete control over their data that are used for public health reporting by configuring a set of FHIR resources and data elements that can be removed with this *FHIR Filter*. The modular and versatile components that form the architecture connect to the EHR and extract data as an electronic case report for public health surveillance without extensive modifications.

Methods

Overview

This project was determined to be a quality improvement project by an automated review tool developed by the Medical University of South Carolina (MUSC) based on the federal definition of research; as such, it was exempted by MUSC from an institutional review board review. We developed and tested the PACER architecture using simulated data first. We then implemented this architecture at MUSC. During this pilot study, we did not enroll any health department, with the intention of testing the PACER architecture in a health care setting first. For these reasons, we simulated the role of the health department. The results were evaluated by MUSC, ensuring that patient data remained with the health system, and only the aggregate findings of the study were reported.

Case Report Trigger Logic

Case reports need a trigger logic, which is a set of criteria that must be satisfied to qualify as a notifiable condition and create an automated case report. The PACER architecture uses incoming ELR to the health department as trigger criteria to generate queries into the EHR. The CSTE has developed standardized reporting definitions for notifiable conditions [19]. The goal of these position statements is to support a standard for case reporting from EHRs or other clinical care information systems. These position statements—which represent documentation and analysis regarding the case definition of the condition—are based on laboratory results, clinical observations (such as diagnosis), or a combination of both. The ELR infrastructure uses the CSTE's position statements to qualify a case and create a report. ELR has seen widespread implementation since it was incentivized by the Centers for Medicare and Medicaid Services Meaningful Use program [20,21]. Electronic laboratory reports are a reliable and widely implemented method for case reporting. However, the major limitation is that electronic laboratory reports do not contain additional and important information needed to understand disease epidemiology.

Design and Testing

Design and testing were performed at the Georgia Tech Research Institute. During this phase, we used ELR messages entering a simulated health department as the initial trigger for PACER (Figure 1). Using the identifier for the patient and the clinical provider in the ELR message, a request for additional case information was sent to the clinical provider (Figure 2). The CQL logical statements were predetermined and provided before the request in the form of a CQL script. This script can be readily updated for the new report logic. A PACER server within the clinical provider's network received and managed the query, as depicted in Figure 3. A CQL repository contained predetermined logical expressions as a CQL query syntax based on the typical request of the health department. On the basis of this CQL, a FHIR application programming interface (API) queried a FHIR server connected to the provider's server, which is, in turn, connected to the EHR. This CQL engine can also integrate the standard and local terminologies. As not all FHIR systems adhere to the same standard codes, PACER was designed to have local codes and queries against the FHIR server. However, in the previous pilot study [18], it was found that this mechanism extracted a lot of redundant data not relevant to that particular reportable condition. In the current architecture, PACER includes a FHIR-based translation API service between the standard and local codes. Clinical providers can construct the local to standard mapping information in a comma-separated values format and load it to the translation API service, which can then be used to translate standard code to local code and vice versa during a CQL over FHIR operation. We also designed a *FHIR Filter*—an API service—which filters FHIR resources. This ensured that only those data elements predetermined by the provider were retrieved for the case report, and all irrelevant data, such as clinical observations and treatment information from unrelated clinical visits, were removed. This mechanism (Figure 2) ensured that the provider systems have complete control over the data extracted from the EHR and are consistent with the case reporting requirements. Filtered information, such as demographics, clinical observations, and treatment information, were retrieved from the EHR and then sent back to the simulated health department as an electronic case report, where it was received and stored in the eCR repository.

This process could potentially be replicated for any other notifiable conditions listed in the CSTE position statements or other similar standards. The PACER server is placed within the provider's network firewall and connects to the FHIR servers of the EHR. This architecture ensures that control over the data extraction and filtering mechanism is entrusted to the providers themselves. Public health departments may only request and receive reports using preapproved queries, thereby avoiding the risk of data overreach or gathering information that is extraneous to the notifiable condition. PACER can be deployed on any internet-enabled server, either on premise or on the cloud. PACER can connect to multiple EHRs simultaneously and can extract case-related information from them.

Figure 1. Electronic laboratory reports sent from laboratories are ingested by health departments and serve as the trigger for PACER electronic case reports. CSTE: Council for State and Territorial Epidemiologists; CQL: Clinical Quality Language; eCR: electronic case report; ELR: electronic laboratory reporting; HL7: health level 7; MLLP: Minimum Lower Level Protocol; PACER: Public Health Automated Case Event Reporting.

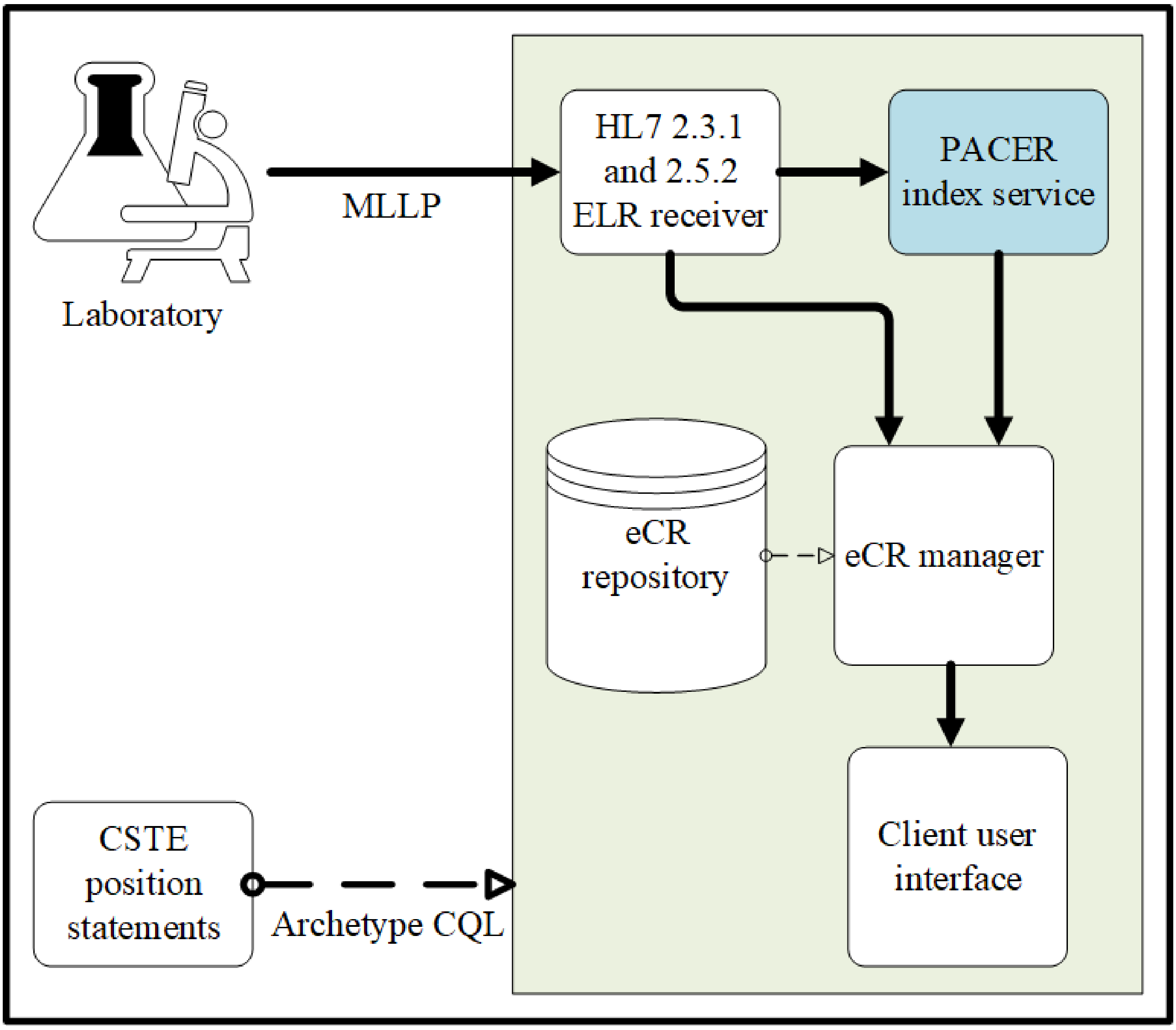

Laboratory
MLLP
HL7 2.3.1 and 2.5.2 ELR receiver
PACER index service
eCR repository
eCR manager
Client user interface
CSTE position statements
Archetype CQL

Figure 2. Detailed architecture of PACER. API: application programming interface; CQL: Clinical Quality Language; eCR: electronic case report; EHR: electronic health record; ELR: electronic laboratory reporting; FHIR: Fast Health Interoperability Resources; HL7: health level 7; OAuth2: OAuth 2.0 Authorization Framework; PACER: Public Health Automated Case Event Reporting.

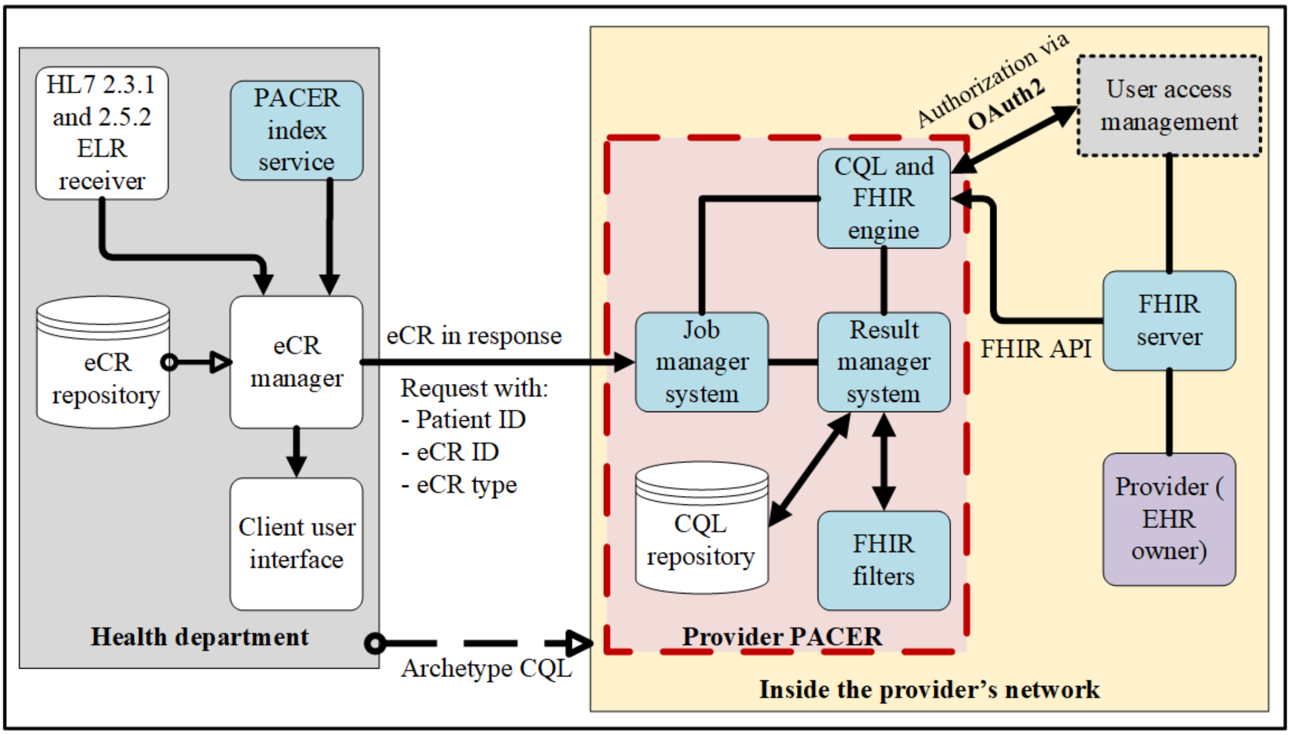
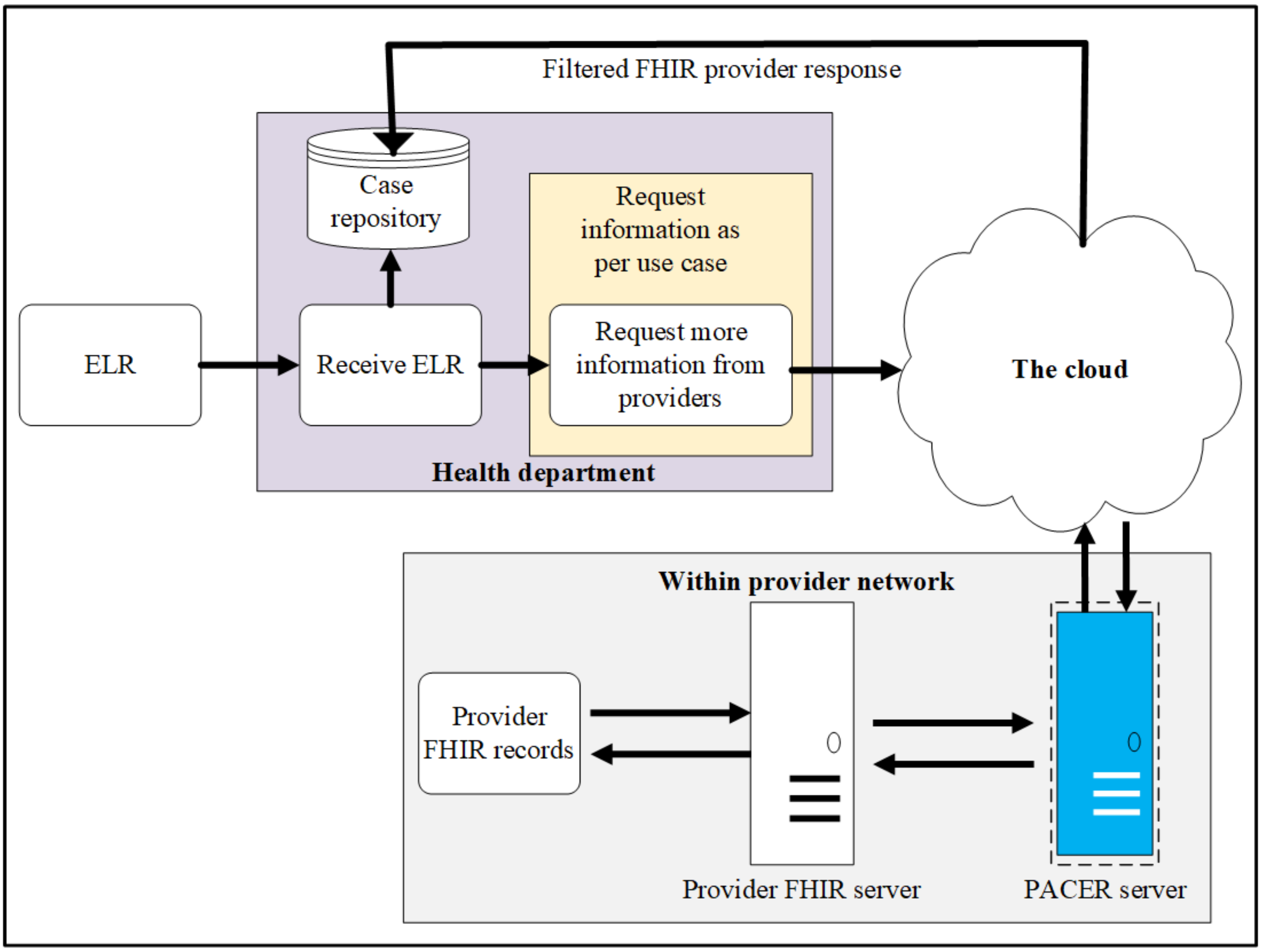

HL7 2.3.1 and 2.5.2 ELR receiver
PACER index service
eCR repository
eCR manager
Client user interface
Health department
eCR in response
Request with:
- Patient ID
- eCR ID
- eCR type
Archetype CQL
Authorization via OAuth2
User access management
CQL and FHIR engine
Job manager system
Result manager system
CQL repository
FHIR filters
FHIR API
FHIR server
Provider (EHR owner)
Provider PACER
Inside the provider's network

Figure 3. High-level architecture of PACER. ELR: electronic laboratory reporting; FHIR: Fast Health Interoperability Resources; PACER: Public Health Automated Case Event Reporting.

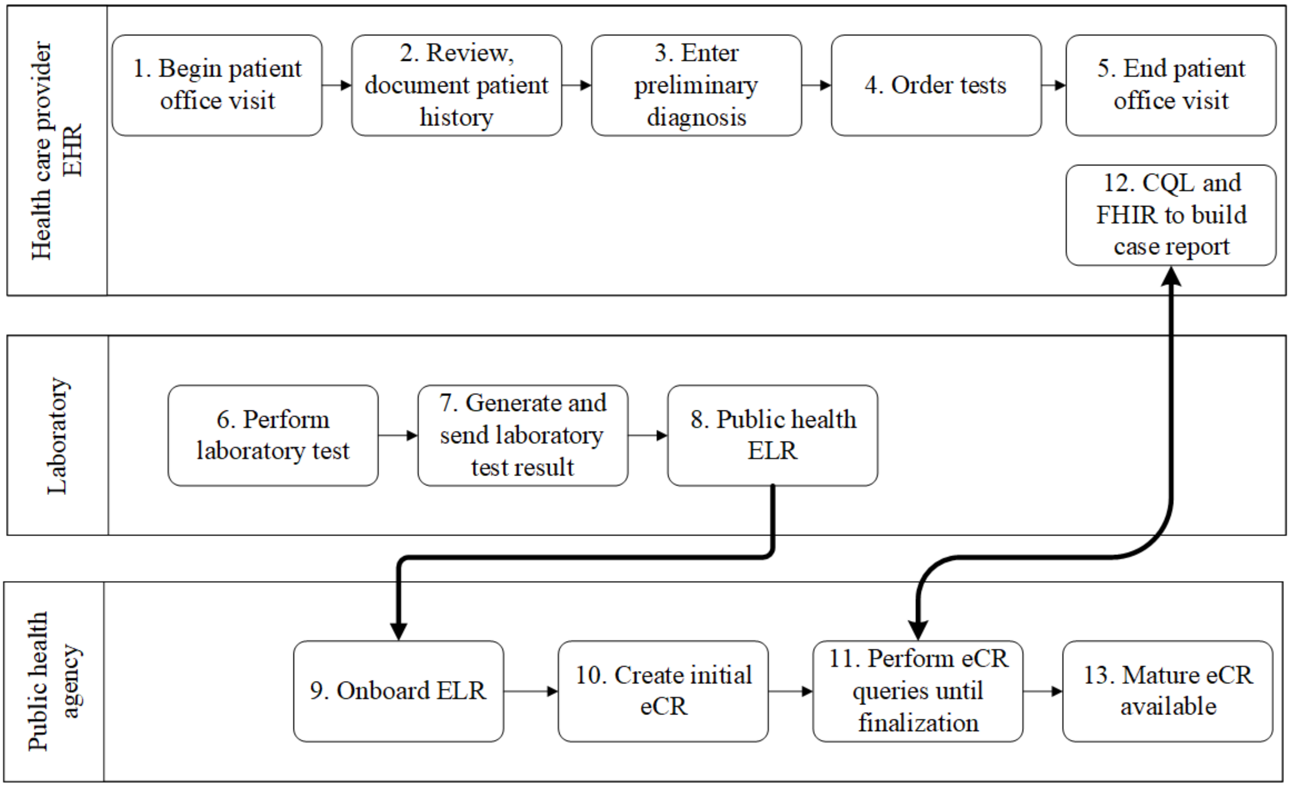

Filtered FHIR provider response
Case repository
Request information as per use case
ELR
Receive ELR
Request more information from providers
Health department
The cloud
Within provider network
Provider FHIR records
Provider FHIR server
PACER server

Pilot Deployment

We deployed PACER at the MUSC Health System by using the architecture shown in Figure 3. We continued to use the Georgia Tech Research Institute–simulated public health department (GTHD) server for the pilot study to ensure that all components were working. The MUSC configured their gateway proxy server to accept API calls from the GTHD server. The CSTE position statements for chlamydia [22] were translated into a CQL query syntax. This incoming positive chlamydia electronic laboratory report triggered an eCR request from the simulated public health department server.

First, full end-to-end testing was performed using synthetic data in the form of ELR sent to the GTHD server, which triggered the eCR manager (Figure 1) to send a request to the MUSC PACER server endpoint configured to use the synthetic FHIR data (Figure 2). From the ELR message, GTHD was able to send the PACER CQL request and successfully receive information to be incorporated into the initial case report, including information unavailable in the original ELR HL7 v2 message. This successfully demonstrated the workflow in Figure 4 from steps 7 to 13. After connectivity testing was completed, PACER was reconfigured to connect to the MUSC data assets. MUSC maintains a research FHIR server (SmileCDR) that contains a subset of its Epic EHR and supports external queries for FHIR resources. This server was used to respond to PACER queries. The outgoing connection was turned off to maintain privacy and confidentiality. Instead, the request was sent to PACER internally. The resulting electronic case reports were stored in a local MUSC environment and were analyzed locally.

Figure 4. Laboratory, public health agency, and health care provider workflow to generate case reports using Public Health Automated Case Event Reporting. CQL: Clinical Quality Language; eCR: electronic case report; EHR: electronic health record; ELR: electronic laboratory reporting; FHIR: Fast Health Interoperability Resources.


Health care provider EHR
1. Begin patient office visit
2. Review, document patient history
3. Enter preliminary diagnosis
4. Order tests
5. End patient office visit
12. CQL and FHIR to build case report
Laboratory
6. Perform laboratory test
7. Generate and send laboratory test result
8. Public health ELR
Public health agency
9. Onboard ELR
10. Create initial eCR
11. Perform eCR queries until finalization
13. Mature eCR available

Evaluation Metrics

We evaluated the data elements extracted by PACER in the form of eCR, triggered by confirmed positive chlamydia laboratory results over an 8-week study period. The CQL query used in the pilot study was designed to obtain relevant data on patient diagnoses, symptoms, medications, and demographics of patients with chlamydia. We chose chlamydia because it was the most commonly reported condition, with more than 1.7 million cases reported in 2018 [23]. All evaluations and data analyses were performed at MUSC, the health system where PACER was deployed, and only the aggregate findings were reported by MUSC staff. The number of actual chlamydia cases and the number of electronic case reports created were compared to determine the success of the pilot. The completeness of case reports was also evaluated by assessing the presence of diagnoses, symptoms, medications, and demographics in each electronic case report. These same data were evaluated to test whether data relevant only to that particular event of chlamydia diagnosis were filtered by PACER. We also evaluated the type of data elements that PACER eCR extracts compared with ELR. The project was evaluated based on the accessible data from the MUSC FHIR server.

Results

By using 8 weeks of laboratory data, 117 positive chlamydia laboratory tests triggered an electronic case report request through the PACER system deployed at MUSC. PACER successfully created 100% (117/117) of the electronic case reports containing data elements specified in the CQL. As shown in Table 1, PACER was able to obtain all key data elements for the majority of patients. In addition to routine demographic data, PACER was able to obtain the clinical diagnosis of 64.9% (76/117) of patients, treatment medications for chlamydia in 72.6% (85/117) of patients, and information on sexually transmitted infection symptoms in 70.9% (83/117) of patients.

Table 1. Patient data retrieved from electronic case reports in electronic health records (N=117).

Data elements	Case reports that contained the data elements, n (%)
Demographics	116 (99.1)
Medications	85 (72.6)
Symptoms	83 (70.9)
Diagnoses	76 (65)

Table 2 provides a perspective on the types of data available in ELR and the PACER eCR. As noted, electronic laboratory reports report laboratory results but with limited demographic data and no information on conditions, medications, or symptoms. Electronic laboratory reports report confirmed positive laboratory findings, which easily met the reporting criteria, and can be successfully used as a trigger for eCR. Complementing the electronic laboratory report with PACER enhanced case reports by extracting critical additional data elements, including medications, symptoms, and clinical diagnoses.

Table 2. Comparison between the case reports from ELR[a] and PACER[b].

Data elements	Case report from ELR	Case report from PACER
Patient demographics	Limited—as available in the laboratory order	Nearly complete if allowed by providers
Laboratory results	Available	Available
Conditions	Not available	Available
Medications	Not available	Available
Extensibility	Limited by ELR specification	Can be extended by defining additional resources in CQL[c]

[a] ELR: electronic laboratory reporting.

[b] PACER: Public Health Automated Case Event Reporting.

[c] CQL: Clinical Quality Language.

Discussion

Principal Findings

We piloted the use of an electronic case report architecture combining the widespread adoption and reliability of ELR with the robustness of EHR data through a system leveraging the rapidly emerging FHIR standard. The CQL and permissions for data access can change dynamically during outbreaks. Thus, an important advantage of the PACER design is its ability to respond to the evolving needs of clinical data for emerging infectious diseases such as COVID-19. This architecture was able to provide additional sexually transmitted infection–relevant clinical data on diagnoses, symptoms, and medications in 65% (76/117), 70.9% (83/117), and 72.6% (85/117) of patients, respectively. Moreover, once the initial queries were approved, these data were captured without human intervention, minimizing the burden of manual chart review. The FHIR query can only bring data available in the EHRs at the time the query reaches the EHR. There could be other factors responsible for the unavailability of the EHR data, such as laboratory tests stored in local codes and not LOINC. Some of this could be mitigated by sending FHIR queries on certain periodic intervals, such as treatment information, which may only be available after a certain time lag post diagnosis. PACER clearly demonstrates advantages in enhancing electronic laboratory reports and C-CDA–based electronic case reports. PACER can extract important demographic, diagnoses, treatment, and other relevant case information that the electronic laboratory reports, by virtue of being reported out of laboratories, will never contain.

C-CDA–based eCR is triggered at a point in the continuum of clinical care [24], such as the initial visit where the case definition is met. Some local health jurisdictions mandate clinical providers to report cases within a definite period of the case being known or diagnosed. Typically, treatment information and other relevant data generated chronologically along the continuum of care will not be available at the time the electronic case report is generated. The PACER architecture remedies these issues. The ELR trigger satisfies the time-bound reporting mandate imposed by some jurisdictions. As ELR triggers FHIR queries at points that are chronologically further along the continuum of care, the PACER architecture can extract such data that may not be available at the time a case report is triggered. In addition, although the C-CDA format extracts a case report in its entirety every time it is triggered [24], the FHIR format allows the query for specific data points. Subsequent queries can be triggered at any time, seeking specific information, such as treatment or follow-up data from EHRs. Within the PACER architecture, the eCR manager handles this request, and multiple information requests can be queried at any point along the care continuum.

In the pilot of an earlier iteration of PACER, we discovered that the architecture extracted every record of the individual, including information that was not relevant to the reportable disease [18]. In this iteration, we added filters to the provider server to ensure that only relevant (and predetermined) information related to the reportable disease is extracted. The filtering of data within a FHIR resource is needed when the FHIR resource contains more data elements than required for public health case reporting. In conjunction with the CQL logic (predetermined between the clinical provider and health departments), this design prevented the release of any unrelated information but allowed for the secure and free flow of information for public health. PACER is an open-source architecture that is maintained on GitHub and is freely accessible by the public. We plan to collaborate with clinical providers, EHR service providers, and local and state health departments to make PACER more robust and easily accessible. We also continue to expand the use of PACER architecture for other public health uses.

Limitations

There are several challenges in the current environment of PACER. First, the design of the clinical queries using CQL relies on understanding the local clinical codes used within the health systems, if applicable. To maintain the quality of eCR outputs over different provider systems, it is necessary to standardize to common code systems at the CQL level while enabling mapping at the individual provider level. This standardization can be facilitated by maintaining value sets that include medical codes used in provider systems and maps from local medical codes to standard codes. For the pilot study, we found that some local medical code mappings were missing and required additional manual effort to finalize the value sets. Complete and accurate mapping is a perennial challenge, and in some contexts, it is worth sacrificing *edge case* codes (ie, rarely used nonstandard codes) to create systems that are portable and generalizable. We also expect the recent formalization of US Core Data for Interoperability to have a positive impact on the ability to retrieve standardized concepts going forward.

The other major consideration of this study is the adoption of FHIR itself. FHIR are rapidly emerging, with more than 80% of US hospitals using a FHIR-capable EHR. Even then, the FHIR capability is not the same as the FHIR capacity. One of the limitations of this work was the use of a dedicated FHIR server that contains only a subset of the data available in the MUSC EHR. However, health systems are still working through the governance processes for how FHIR will be used in their environments, and thus, the coordination between health systems' information technology departments and public health will be essential for successful automated reporting, as demonstrated here. For mitigating these challenges, PACER has been designed to be (1) highly protective of health system data with privacy controls and (2) flexible enough to support other health system operations and data query needs that may fall outside of public health reporting. This represents a possible *carrot* to health systems considering the implementation of a system but are cautious regarding dedicating information technology resources to a single use-case effort.

Conclusions

Enhancing electronic laboratory reports with FHIR-based EHR data using an automated system such as PACER can expand access to relevant clinical data for public health reporting and decision-making. This method can extract important data to understand public health epidemiology and, in turn, inform policy and strategies to prevent and control sexually transmitted diseases in general and potentially aid in response to emerging infectious disease threats. Using ELR—or even eCR—as a trigger for FHIR-based queries that create additional electronic case reports with other relevant information is an adaptable and modular approach to enhancing case reporting. Clinical providers will benefit from this platform by automating the reporting of notifiable diseases and eliminating the burden of manual reporting. More importantly, clinical providers will have complete control of what data are reported to public health agencies. This is achieved through FHIR-based filters and can be further modified by the clinical providers at any time. PACER is EHR agnostic and is installed and invoked in a centralized manner, eliminating the need for multiple installations at every instance of an EHR. This approach not only alleviates the burden of manual reporting from providers but also public health departments get timely, accurate (from the EHR—the source of the clinical interaction), and more complete and detailed information on a case. As the type of information that can be extracted is predetermined (using CQL-based logic), any type of information can be retrieved as long as both the clinical provider and public health agencies agree on them. This will lead to efficient case investigations with essential information that is automatically populated. Traditionally, public health has struggled with getting information, such as treatment, medications, and symptoms. This approach provided additional information that will help public health better understand disease epidemiology. This architecture can be implemented for any other notifiable condition that has CSTE position statements translated into a logical format or even used for distributed data collection for an emerging infection such as COVID-19. This design emphasizes giving complete control of patient data to clinical providers, with the ability to filter what information is sent to public health agencies as part of case reports for notifiable diseases. This architecture is a step toward bridging public health with health care systems and can be used to develop a *query-based* system to gain important and relevant information related to public health priority. Future work will include expanding the pilot study to multiple facilities in local public health jurisdictions and engaging with the health systems within those jurisdictions to set up PACER. In addition, we are currently expanding PACER to build a FHIR-based syphilis registry.

Acknowledgments

The development and conceptualization of PACER were funded by the Centers for Disease Control and Prevention under contract D8954 at the Georgia Tech Research Institute. The findings and conclusions in this report are of the authors alone and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Conflicts of Interest

None declared.

References

1. Thacker SB, Berkelman RL. Public health surveillance in the United States. Epidemiol Rev 1988;10:164-190. [doi: 10.1093/oxfordjournals.epirev.a036021] [Medline: 3066626]
2. Wharton M, Chorba TL, Vogt RL, Morse DL, Buehler JW. Case definitions for public health surveillance. MMWR Recomm Rep 1990 Oct 19;39(RR-13):1-43 [FREE Full text] [Medline: 2122225]
3. Centers for Disease Control and Prevention. Case definitions for infectious conditions under public health surveillance. Centers for Disease Control and Prevention. MMWR Recomm Rep 1997 May 02;46(RR-10):1-55 [FREE Full text] [Medline: 9148133]
4. Birkhead GS, Klompas M, Shah NR. Uses of electronic health records for public health surveillance to advance public health. Annu Rev Public Health 2015 Mar 18;36:345-359. [doi: 10.1146/annurev-publhealth-031914-122747] [Medline: 25581157]
5. Mac Kenzie WR, Davidson AJ, Wiesenthal A, Engel JP, Turner K, Conn L, et al. The Promise of Electronic Case Reporting. Public Health Rep 2016 Nov;131(6):742-746 [FREE Full text] [doi: 10.1177/0033354916670871] [Medline: 28123218]
6. HL7 International. HL7 Version 2 Product Suite. URL: https://www.hl7.org/implement/standards/product_brief.cfm?product_id=185 [accessed 2020-05-05]
7. Dixon BE, Gibson PJ, Grannis SJ. Estimating increased electronic laboratory reporting volumes for meaningful use: implications for the public health workforce. Online J Public Health Inform 2014;5(3):225 [FREE Full text] [doi: 10.5210/ojphi.v5i3.4939] [Medline: 24678378]
8. HL7 International. C-CDA (HL7 CDA® R2 Implementation Guide: Consolidated CDA Templates for Clinical Notes - US Realm). URL: http://www.hl7.org/implement/standards/product_brief.cfm?product_id=492 [accessed 2020-05-05]
9. Weston EJ, Workowski K, Torrone E, Weinstock H, Stenger MR. Adherence to CDC Recommendations for the Treatment of Uncomplicated Gonorrhea - STD Surveillance Network, United States, 2016. MMWR Morb Mortal Wkly Rep 2018 Apr 27;67(16):473-476 [FREE Full text] [doi: 10.15585/mmwr.mm6716a4] [Medline: 29698384]
10. Thacker SB, Qualters JR, Lee LM, Centers for Disease ControlPrevention. Public health surveillance in the United States: evolution and challenges. MMWR Suppl 2012 Jul 27;61(3):3-9. [Medline: 22832990]
11. Patient Safety: Achieving a New Standard for Care. Aspden P, Corrigan JM, Wolcott J, Erickson SM. editors. Washington, DC: National Academies Press; 2004.
12. Pfaff ER, Champion J, Bradford RL, Clark M, Xu H, Fecho K, et al. Fast Healthcare Interoperability Resources (FHIR) as a Meta Model to Integrate Common Data Models: Development of a Tool and Quantitative Validation Study. JMIR Med Inform 2019 Oct 16;7(4):e15199 [FREE Full text] [doi: 10.2196/15199] [Medline: 31621639]
13. LOINC. Regenstrief Institute. URL: https://loinc.org [accessed 2020-05-05]
14. SNOMED International. SNOMED CT. URL: http://www.snomed.org [accessed 2020-05-05]
15. National Library of Medicine. ICD10CM. URL: https://www.nlm.nih.gov/research/umls/sourcereleasedocs/current/ICD10CM/index.html [accessed 2020-05-05]
16. FHIR overview. HL7 International. URL: https://www.hl7.org/fhir/overview.html [accessed 2019-08-01]
17. HL7 International. Clinical Quality Language (CQL). URL: https://cql.hl7.org/ [accessed 2020-05-05]
18. Dixon BE, Taylor DE, Choi M, Riley M, Schneider T, Duke J. Integration of FHIR to Facilitate Electronic Case Reporting: Results from a Pilot Study. Stud Health Technol Inform 2019 Aug 21;264:940-944. [doi: 10.3233/SHTI190362] [Medline: 31438062]
19. Council for State and Territorial Epidemiologists (CSTE). CSTE. Position Statement Archive. URL: https://www.cste.org/page/PositionStatements [accessed 2019-08-20]
20. EHR Incentive Program. Eligible Hospital and Critical Access Hospital Meaningful Use Menu Set Measures, Stage 1. URL: https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/downloads/9_Reportable_Lab_Results_to_Public_Health_Agencies.pdf [accessed 2019-08-21]
21. EHR Incentive Program. Eligible Hospital and Critical Access Hospital Meaningful Use Core Measures, Stage 2. URL: https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/downloads/Stage2_HospitalCore_14_SubLabResults.pdf [accessed 2019-08-21]
22. Council of State and Territorial Epidemiologists (CSTE). Public Health Reporting and National Notification for Infection Caused by Chlamydia. URL: https://cdn.ymaws.com/www.cste.org/resource/resmgr/PS/09-ID-08.pdf [accessed 2019-08-21]
23. Sexually Transmitted Disease Surveillance 2018. Atlanta, USA: U.S. Department of Health and Human Services; Oct 2019.

24. Whipple A, Jackson J, Ridderhoff J, Nakashima AK. Piloting Electronic Case Reporting for Improved Surveillance of Sexually Transmitted Diseases in Utah. Online J Public Health Inform 2019;11(2):e7 [FREE Full text] [doi: 10.5210/ojphi.v11i2.9733] [Medline: 31632601]

Abbreviations

API: application programming interface
C-CDA: Consolidated Clinical Document Architecture
CQL: Clinical Quality Language
CSTE: Council for State and Territorial Epidemiologists
eCR: electronic case reporting
EHR: electronic health record
ELR: electronic laboratory reporting
FHIR: Fast Health Interoperability Resources
GTHD: Georgia Tech Research Institute–simulated public health department
HL7: health level 7
LOINC: Logical Observation Identifiers Names and Codes
MUSC: Medical University of South Carolina
PACER: Public Health Automated Case Event Reporting

Edited by R Kukafka; submitted 10.12.20; peer-reviewed by N Shah, I Mircheva; comments to author 29.01.21; revised version received 17.02.21; accepted 17.06.21; published 11.08.21.

Please cite as:
Mishra N, Duke J, Karki S, Choi M, Riley M, Ilatovskiy AV, Gorges M, Lenert L
A Modified Public Health Automated Case Event Reporting Platform for Enhancing Electronic Laboratory Reports With Clinical Data: Design and Implementation Study
J Med Internet Res 2021;23(8):e26388
URL: https://www.jmir.org/2021/8/e26388
doi:10.2196/26388
PMID:34383669

©Ninad Mishra, Jon Duke, Saugat Karki, Myung Choi, Michael Riley, Andrey V Ilatovskiy, Marla Gorges, Leslie Lenert. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 11.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Interorganizational Knowledge Sharing to Establish Digital Health Learning Ecosystems: Qualitative Evaluation of a National Digital Health Transformation Program in England

Kathrin Cresswell[1], PhD; Aziz Sheikh[1], MD; Bryony Dean Franklin[2], PhD; Marta Krasuska[1], PhD; Hung The Nguyen[3], PhD; Susan Hinder[3], PhD; Wendy Lane[4], MA; Hajar Mozaffar[5], PhD; Kathy Mason[4], MSc; Sally Eason[4], MSc; Henry Potts[6], PhD; Robin Williams[3], PhD

[1]Usher Institute, The University of Edinburgh, Edinburgh, United Kingdom
[2]School of Pharmacy, University College London, London, United Kingdom
[3]Institute for the Study of Science, Technology and Innovation, The University of Edinburgh, Edinburgh, United Kingdom
[4]National Health Services Arden and Greater East Midlands Commissioning Support Unit, Warwick, United Kingdom
[5]Business School, The University of Edinburgh, Edinburgh, United Kingdom
[6]Institute of Health Informatics, University College London, London, United Kingdom

Corresponding Author:
Kathrin Cresswell, PhD
Usher Institute
The University of Edinburgh
Teviot Place
Edinburgh, EH8 9AG
United Kingdom
Phone: 44 (0)131 651 4151
Email: kathrin.cresswell@ed.ac.uk

Abstract

Background: The English Global Digital Exemplar (GDE) program is one of the first concerted efforts to create a digital health learning ecosystem across a national health service.

Objective: This study aims to explore mechanisms that support or inhibit the exchange of interorganizational digital transformation knowledge.

Methods: We conducted a formative qualitative evaluation of the GDE program. We used semistructured interviews with clinical, technical, and managerial staff; national program managers and network leaders; nonparticipant observations of knowledge transfer activities through attending meetings, workshops, and conferences; and documentary analysis of policy documents. The data were thematically analyzed by drawing on a theory-informed sociotechnical coding framework. We used a mixture of deductive and inductive methods, supported by NVivo software, to facilitate coding.

Results: We conducted 341 one-on-one and 116 group interviews, observed 86 meetings, and analyzed 245 documents from 36 participating provider organizations. We also conducted 51 high-level interviews with policy makers and vendors; performed 77 observations of national meetings, workshops, and conferences; and analyzed 80 national documents. Formal processes put in place by the GDE program to initiate and reinforce knowledge transfer and learning have accelerated the growth of informal knowledge networking and helped establish the foundations of a learning ecosystem. However, formal networks were most effective when supported by informal networking. The benefits of networking were enhanced (and costs reduced) by geographical proximity, shared culture and context, common technological functionality, regional and strategic alignments, and professional agendas.

Conclusions: Knowledge exchange is most effective when sustained through informal networking driven by the mutual benefits of sharing knowledge and convergence between group members in their organizational and technological setting and goals. Policy interventions need to enhance incentives and reduce barriers to sharing across the ecosystem, be flexible in tailoring formal interventions to emerging needs, and promote informal knowledge sharing.

(J Med Internet Res 2021;23(8):e23372) doi:10.2196/23372

KEYWORDS

digital transformation; health system; learning ecosystem

Introduction

Background

Digital transformation is now central to most health system strategies, as governments around the world seek to address the challenges associated with demographic shifts and the need for sustainable care provision to aging populations with complex long-term conditions [1,2]. Although policy makers and implementers generally agree on the potential of health information technology (HIT) to improve safety, quality, and efficiency of care, strategies for procurement, implementation, and optimization vary significantly across settings [3-5]. Large-scale HIT-enabled transformation programs have met with varying success, and there is no agreed strategy on how best to achieve digital transformation at scale [6-9].

Many aspects of digital transformation have been studied [10-12]. However, interorganizational knowledge sharing is a key feature of recent initiatives to promote concerted change across multiple organizations by establishing a learning ecosystem [13,14]. Understanding interorganizational knowledge transfer may help to mitigate risks by avoiding repetition of mistakes, thereby saving money and minimizing potential threats to patient safety and quality of care. Concerted adoption might also reduce inefficiencies of fragmented one-off implementations by encouraging learning across communities of adopters and increasing their influence on system development.

We use the term *learning ecosystem* to refer to interorganizational sharing of technology, knowledge, and know-how to achieve digital transformation (ie, to change technologies and organizations). We differentiate this from the notion of *learning health systems*, which focuses on optimizing the use of clinical and operational data to advance and apply medical research (ie, to advance the clinical cycle or improve care processes) [15]. The learning ecosystem highlights that knowledge and experience of technology adoption and implementation is particularly valuable for members of other organizations contemplating similar digitally enabled transformation (as well as for vendors and policy makers) [16].

Although there have been local examples of attempts to promote digital health–related knowledge exchange, these are often not systematically evaluated and are poorly theorized [17,18]. In contrast, in the commercial sector, a large body of literature explores knowledge transfer between technology vendors and users [19-22]. This work highlights the key role of various intermediaries in bridging gaps, translating, and facilitating information flows between different stakeholder groups [23,24]. In addition to formal organizational links (eg, vendor-hosted user groups), informal networking, driven by the benefits of knowledge transfer, can be particularly important in communicating *sticky* information (information that is hard to acquire and intimately linked to its context of use) [25]. Some papers discuss user-to-user sharing of knowledge, but this focuses mainly on consumer products or open-source applications [26,27].

Objectives

We were commissioned to conduct an independent evaluation of the Global Digital Exemplar (GDE) program. This was the first national digital transformation program designed to facilitate concerted interorganizational knowledge exchange and create a digital health learning ecosystem [28,29]. This program was set up following a national review of national HIT strategy led by Robert Wachter [30], with the National Health Service (NHS) England committing £395 million (US $544 million) to support the development and interorganizational knowledge sharing between a selected group of digitally mature provider organizations. A total of 51 organizations participated in the program, including 33 acute care, 15 mental health, and 3 ambulance provider organizations. The latter 3 were not included in the evaluation along with 9 sites whose launch was delayed, and 3 where organizations merged. As a result, the evaluation covered 36 provider organizations. Some of these were designated leaders who were to implement first (GDEs) and others partnered with GDEs to learn from and follow them (fast followers).

We sought to answer the following research question: How is interorganizational knowledge sharing taking place within the GDE program?

Methods

Setting

The GDE program's attempt to establish a digital health learning ecosystem was accompanied by related national initiatives, including professional training and education [31]. Specific mechanisms to promote interorganizational knowledge transfer included the following:

1. GDE and fast follower pairings: This involved pairing digitally advanced exemplar provider organizations (GDEs) with partner organizations (fast followers) who would follow and learn from GDEs throughout the duration of the program. The rationale for the pairings varied among stakeholders, and no official documentation was available on the issue. Most organizations appeared to choose their own partners. Other pairings were established by external stakeholders. Care settings were paired with each other so that mental health organizations were matched with other mental health organizations, and acute organizations were paired with other acute organizations.
2. Establishing a series of national learning networks to promote knowledge transfer among participating provider organizations and across the wider NHS.
3. Blueprinting: This involved asking all participating provider organizations to produce documents (blueprints) to capture implementation, adoption and optimization experiences.

Design

We conducted an independent, longitudinal, qualitative, formative evaluation of the GDE program, exploring digital transformation and knowledge transfer in participating acute and mental health provider organizations. The study period was from January 2018 to March 2020. The evaluation revolved around two core themes: first digital transformation of sites, and second, the processes of interorganizational knowledge transfer explored in this paper. Some findings have been reported in previous studies [32,33]. Our evaluation reports are available on our website [29].

Methods included a combination of in-depth semistructured one-on-one and group interviews with relevant organizational stakeholders (managers and clinicians), documentary analyses of organizational strategic plans, and ethnographic fieldwork (nonparticipant observations of strategic meetings and site visits) to explore national knowledge networks and linkages between organizations. This allowed insights into local knowledge, organizational context and progress, and formal and informal knowledge transfer mechanisms as experienced by those participating in knowledge transfer activities. We also collected a range of national documents addressing planned knowledge transfer mechanisms, observed national workshops and conferences where knowledge was shared formally and informally, and conducted in-depth interviews with national program managers and system vendors to gain insights into how they planned and participated in knowledge sharing.

Provider organizations were conceptualized as case studies [34]. We conducted in-depth studies of 12 organizations (A-M) and broader case studies of the remaining 24 organizations. In-depth case studies were selected for maximum variation, including different core technological systems, geographical locations, organizational types (acute and mental health), and baseline levels of digital maturity. Detailed methods of our full evaluation, of which the creation of a learning ecosystem was a central theme, are described in our study protocol [35].

Analysis

We combined inductive and deductive methods, drawing on a theory-informed, sociotechnical coding framework [36]. Lead researchers (SH, MK, and HTN) initially analyzed knowledge flows in in-depth case studies and then tested and validated these emerging findings against broader case studies. We then integrated developing narratives with accounts of the wider macroenvironmental landscape from policy, commercial, and independent stakeholders and tested these against case study data [37]. Narrative accounts were produced collectively and through discussions in team meetings, where we paid most attention to emerging tensions and conflicting findings. We have published detailed empirical accounts of the operation of individual learning components initiated through the program elsewhere [32,33]. The current analysis focused on integrating different strands of inquiry relating to knowledge networks across the GDE program and within the wider macroenvironmental context. In doing so, we also identified the roles and effectiveness of various intermediaries. We used the COREQ (Consolidated Criteria for Reporting Qualitative Studies) guidelines [38].

Ethical Approval

This work received institutional ethical approval from the School of Social and Political Science at the University of Edinburgh, UK.

Results

Overview

We conducted 341 one-on-one and 116 group interviews, observed 86 meetings, and analyzed 245 documents from 36 participating provider organizations (Textbox 1; Table 1; Table S1 of Multimedia Appendix 1). We also conducted 51 high-level interviews with policy makers and vendors; 77 observations of national meetings, workshops, and conferences; and analyzed 80 national documents.

Textbox 1. Study data set.

Data Collected in the In-depth Case Study Sites (Global Digital Exemplar [GDE])

- 12 provider organizations
- 8 GDEs: 6 acute and 2 mental health
- 4 fast followers: 3 acute and 1 specialist fast follower
- 224 one-on-one interviews
- 67 group interviews
- 104 documents
- 67 meetings observed

Data Collected in the Broad Case Study Sites

- 24 provider organizations
- 15 GDEs: 10 acute and 5 mental health
- 9 acute fast followers
- 117 one-on-one interviews
- 49 group interviews
- 141 documents
- 19 meetings observed

Data Collected Elsewhere

- 51 high-level interviews with policy makers and vendors
- Nonparticipant observations of 77 national meetings, workshops, and conferences
- 80 documents

Table 1. Description of the sample in the wider Global Digital Exemplar Program landscape (n=48)[a].

Overall	Included in in-depth studies (n=12), n (%)	Included in broader studies (n=24), n (%)	Omitted because of late admission to program (n=9), n (%)	Omitted because fast follower merged with GDE[b] (n=3), n (%)	Total (n=48), n (%)
Overall number of GDEs (excluding ambulance GDEs): 16 acute and 7 mental health	8 (66)	15 (63)	0 (0)	0 (0)	23
Overall number of fast followers: 17 acute and 8 mental health	4 (33)	9 (38)	9 (100)	3 (100)	25

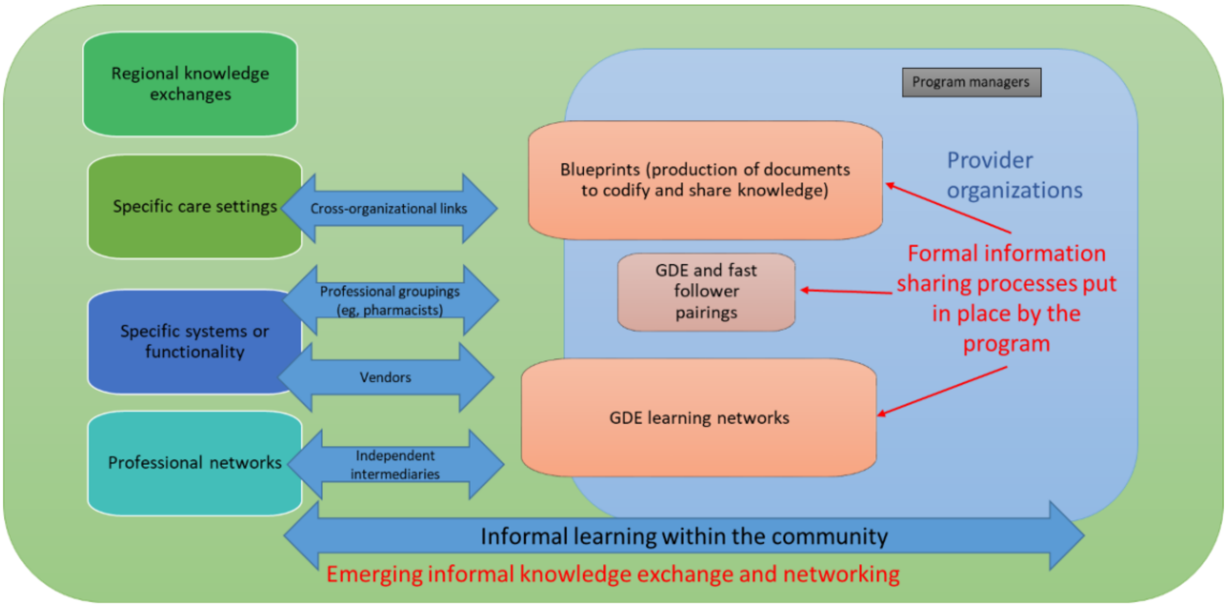
[a] The number of overall Global Digital Exemplars and fast followers differ from those included in our study, as there were some mergers and delays in start dates, which meant that we did not include some provider organizations.

[b] GDE: Global Digital Exemplar.

In our broader sample, 19 pairings of GDEs and fast followers had a common core system, and 15 organizations were in the same local strategic groupings coordinating collaborations of health care organizations and local authorities (including so-called sustainability and transformation partnerships and integrated care systems). These local strategic groupings developed in parallel with the program. In our 12 in-depth case studies, 6 pairings were located in the same local strategic grouping, and 10 had the same core system as their fast follower.

Figure 1 illustrates the emerging formal and informal learning and knowledge exchange processes, knowledge exchange forms, and key intermediaries in the program. We use the term *formal* to describe knowledge exchange processes resulting directly from planned program activities, including those emerging from GDE and fast follower relationships, blueprinting of documents, and program learning networks. We use the term *informal* to describe emerging knowledge exchange processes either as an unanticipated, indirect consequence of these activities or as unrelated activities.

Figure 1. Formal and informal learning and knowledge exchange processes in the Global Digital Exemplar program. GDE: Global Digital Exemplar.


Regional knowledge exchanges
Specific care settings
Cross-organizational links
Professional groupings (eg, pharmacists)
Specific systems or functionality
Vendors
Professional networks
Independent intermediaries
Program managers
Blueprints (production of documents to codify and share knowledge)
Provider organizations
GDE and fast follower pairings
Formal information sharing processes put in place by the program
GDE learning networks
Informal learning within the community
Emerging informal knowledge exchange and networking

Overall, our work suggests that GDE initiatives, coupled with the broader impetus generated by the program, have promoted a burgeoning learning culture across digitally engaged provider organizations and GDE and fast follower pairs, with increased sharing of knowledge and experience. All but 5 provider organizations in our sample described involvement in networking activities, sharing knowledge and experience, and learning from others. We also observed some evidence of the emergence of a learning ethos in the NHS reinforced by these processes:

> *...[W]e're starting to share what we're doing, in a demonstrable way, and start to see it, and it was quite powerful.* [Site 14, nonclinical digital leader, broader case study]
>
> *[Provider organization] had spent about a year building pediatric medicines. And they said here, you can have it. So that's a year's work, that's non-trivial. They just simply gave it to us. Now would that have happened two years ago? Three years ago?...So there are people sharing things of real value, real cost, real-time...which is excellent. So are we creating new knowledge by that, I'm not sure. Are we sharing and optimizing that knowledge? Very definitely.* [Site L, chief information officer, in-depth case study]

Evolving Formal Processes to Promote a National Digital Health Learning Ecosystem

Evolving Formal Processes

Program managers implemented several linked formal initiatives to facilitate knowledge transfer within tight time frames. Formal mechanisms were encouraged and strongly supported by the burgeoning of informal networking and sharing of knowledge and experience. These developments led to changes in the strategic focus of the program. In particular, the strategy associated with the production and distribution of blueprints evolved to become a key component of the learning ecosystem. The blueprinting process changed as the user community (provider organizations) became actively engaged in developing the mechanisms for their production, distribution, and use. Blueprints were initially conceived as repositories of the extensive information needed for the rapid procurement and implementation of validated technologies that could then be widely disseminated. However, provider organizations found them useful in unanticipated ways—as an initial introduction to a topic and as a way to identify and make contact with people involved in implementations—leading to email exchanges, phone calls, and site visits. Thus, blueprinting changed from an activity of capturing digital transformation knowledge in artifacts to a means of facilitating informal networking:

> *[Blueprinting]'s supposed to be not just about taking and adopting, it's to open up conversations.* [Site 8, nonclinical digital leader, broader case study]

The evolving blueprinting concept also saw a relaunched web-based platform, radically reconceptualizing blueprints as *a structured collection of knowledge assets and associated methodology for using them* [39]. It largely overtook centrally driven GDE learning networks; however, where professional groups drove learning networks and where networks tackled specific functionality, these were very successful in attracting and sustaining participation and became national communities of practice. Occupational groupings that aligned their professional interests with enhancing practice through digital transformation were particularly successful. For example, pharmacists were actively involved in knowledge networks around hospital electronic prescribing and medicines administration systems:

> *...[A]ll the GDE groups that work on prescribing, we're having monthly phone calls and meetings* [Site H, senior manager, in-depth case study]

We observed that national activities in many instances helped to initiate and sustain informal networking. Where informal networking emerged, it maximized the effectiveness of formal interorganizational knowledge transfer processes and ensured their sustainability:

> *...[N]othing is really very formal any more, they will pick up the phone and phone [other GDE] and ask how they are doing it. So, it's those informal relationships that I think are really beneficial.* [Site B, GDE program staff, in-depth case study]

Contextual Factors Influencing Informal Knowledge Sharing

Contextual Factors in Provider Organizations

Although informal processes constituted a large and effective part of knowledge transfer and networking, these varied significantly among participating provider organizations. Analyzing these differences provides insights into facilitators and barriers to knowledge transfer.

We found that the adoption of a common core system (such as for electronic health records and hospital electronic prescribing and medicines administration systems), prior relationships, geographical proximity, and regional alignment were, in most instances, beneficial for knowledge sharing and networking by reducing the costs of establishing and maintaining interactions.

Participants also frequently mentioned having similar organizational ethos and culture and similar (or the same) patient populations as facilitators:

> *...[W]e're a similar size as [organization] with a similar footprint of patients with similar economic and geographical pressures, so that's really helpful.* [Site C, chief nursing informatics officer, in-depth case study]

Conversely, differences between organizations in culture, patient populations, and needs were seen as barriers to knowledge sharing.

Common Challenges and Technological Functionality

Similarities between organizational settings reduced learning costs and increased the relevance and benefits of knowledge exchange. The common challenges faced by specific care settings were also facilitators of informal interorganizational networking and knowledge transfer. For instance, we observed productive knowledge exchanges among mental health providers. These shared specific needs and purposes (that might be overlooked by larger acute hospitals) and began to organize informal collaboration.

Common technological functionality was a key facilitator, as organizations with the same vendor often faced similar challenges and sharing of lessons could contribute to avoiding repeating mistakes. There was also scope to transfer detailed elements of system configuration, removing the need to replicate onerous coding work and speeding up implementation:

> *...[T]hat has been happening...the knowledge sharing, especially in those organizations with similar systems, oh, you've just done that, so we'll go and look at it, you've done that, we'll take this.* [Site G, senior manager, in-depth case study]
>
> *Clinically, I think it's fantastic, and organizationally and operationally with [GDE], because you've got the same system and we're taking a lot of their content that they're developing and then we copy it.* [Site B, chief medicines information officer, in-depth case study]

The GDE program further encouraged links between users and vendors, including the development of user groups around major system suppliers. In some instances, organizations reported increased leverage over system vendors and joint procurement:

> *...[W]orking with other GDEs has...given us a bigger voice to talk to suppliers, it's given us an opportunity to introduce new people into the market, and then share that experience with others.* [Site F, chief information officer, in-depth case study]

Reputational Benefits and Competition

The reputational benefits of GDE membership were an important motivator for knowledge sharing, but provider organizations were in some respects competing for status and resources, which inhibited knowledge sharing. In particular, some fast follower organizations were unhappy to be designated as *followers*, especially where they felt they possessed, or would soon attain, greater capability than their GDE:

> *I don't call this fast follower I like the word partner...I think that some of the work that we're doing we're leading rather than following our GDE.* [Site B, information management and technology manager, in-depth case study]

One organization was concerned about reputational risk if their partner performed poorly:

> *I think people are worried about reputational damage. So, if the [provider organization] that you were partnered with would never ever get to a position where you were, is that a failure on the mentoring a [provider organization], or is it a failure with the [provider organization] trying to catch up?* [Site A, information technology manager, in-depth case study]

Some GDEs were seeking recognition as the most digitally mature provider organization in the country. Although under some circumstances, organizational status conflicts had inhibited knowledge sharing (eg, where there was a history of local competition between neighboring organizations), these were exceptions to a broader pattern whereby a culture of sharing prevailed.

Mediators Facilitating Knowledge Transfer Across the Wider Health System

Some stakeholders acted as knowledge exchange mediators, extracting and collating lessons from particular implementations for wider applications. Here, a range of interorganizational networks facilitated knowledge exchanges between provider organizations. These included system vendors who coordinated networking among national organizations with the same system (eg, through user groups and pilot site visits, connecting key individuals to work together across organizations) and promoted connections with international organizations with the same system:

> *[Place in the United States] was one we met through [vendor], because they're a [vendor] client, and*

> *[name], who's their Chief Clinical Information Officer, came here, and again we've kept in touch with them.* [Site 19, chief information officer, broader case study]

Professional networks also played an important role. These allowed members with a common interest to get together, and to exchange ideas, challenges, and lessons learned in a neutral space.

Moreover, we observed the development of specialist digital transformation managerial communities that facilitated informal networking. An example here included the formation of an informal national network of chief clinical information officers and a range of web-based and face-to-face networking activities organized by an independent community of digital health professionals [40]:

> *There's an outfit called Digital Health Networks...and they run a series of forums...it's an online community that's growing all the time, and is exchanging ideas very productively.* [Site C, clinical digital lead, in-depth case study]

Another example was the NHS Digital Academy, a national program to develop digital health leadership capabilities in the NHS [31]. During our data collection period, 50 participants from 29 different GDE provider organizations studied at the NHS Digital Academy:

> *...[T]he Digital Academy has really shown that it's phenomenally important...we've had loads of conversations, over dinner and things...about what they're doing, what we're doing...and, actually, that's been really beneficial because otherwise we probably wouldn't have found time to have those conversations.* [Site C, clinical digital lead, in-depth case study]

Relative Costs and Efforts Associated With Knowledge Transfer

The mutual benefits of shared learning and an ethos of public health benefit facilitated emerging small-scale exchanges. The biggest barrier to knowledge transfer cited in our sample was competing demands on participants' time, particularly given the priorities for health professionals to provide day-to-day care:

> *...[Knowledge transfer is] one of those things that you need to make time for and we're all really busy in our day-to-day roles...* [Site D, chief nursing informatics officer, in-depth case study]

Knowledge sharing activities were particularly burdensome for organizations (mainly GDEs) that were perceived as national leaders and, therefore, had many requests from a range of other organizations to share knowledge. Those seeking to establish themselves as national leaders expressed concern that moving forward as a group of organizations could slow down processes such as procurement and thereby hold back their development:

> *...[T]hat's just the difficulty of moving together as a group of organizations, even though we do work very well as a unit. It's those sorts of things where there are more complications in terms of procurement and contracting and so on and so forth.* [Site F, chief information officer, in-depth case study]

Knowledge sharing through informal networking demands people's time and offers fewer obvious opportunities for economies of scale than, for example, circulating documents. There were some concerns that the cost of networking would threaten the sustainability of sharing activities.

Individuals and organizations benefited from learning by receiving information. They could also experience reputational benefits that could improve their status and strengthen individual expert careers. Networking and knowledge transfer were enhanced when the learning costs were minimized and the benefits maximized. However, issues emerged where there was asymmetry between knowledge provision and knowledge receipt for organizations making this informal mutuality difficult to sustain. For example, this was an issue where provider organizations engaged with large numbers of adopters and where knowledge transfer took a lot of resources.

Nationally organized activities mitigated barriers to an extent by reducing the cost of knowledge transfer to provider organizations. Different types of national interventions played a catalytic role. Critical factors included stimulating discussion topics and shaping agendas, setting up webinars and knowledge transfer work, and curating artifacts for sharing:

> *...[W]e've had the capacity to go out and talk to other organizations across the UK which we've done...and the project team have the capacity and the ability to do that. We would never have been able to do that pre-GDE.* [Site E, GDE program staff, in-depth case study]

Discussion

Principal Findings

Our exploration of interorganizational knowledge transfer in the GDE program shows that the program has made a major contribution to the current upsurge in knowledge transfer across the NHS. The combination of formal learning mechanisms and processes to initiate a national digital health learning ecosystem promoted systemic learning; however, it was most successful when supported by informal networks. Formal knowledge transfer mechanisms did not necessarily work in a planned manner. They evolved over time and prompted a dramatic growth in informal learning among organizations and specialist communities.

Strengths and Limitations

We conducted a national formative evaluation of a first-of-type digitally enabled national transformation program and collected a large qualitative data set from a range of settings and data sources. Our research design, combining in-depth with broader data collection, allowed us to balance the depth and breadth of insights. We achieved this by analyzing change processes and mechanisms of knowledge transfer in detailed studies of selected provider organizations, while testing these emerging findings in a wider range of settings and placing them within the national context of the program. In doing so, we gained rich insights into how knowledge transfer took place in an evolving interorganizational learning ecosystem.

However, as this work is based on a national qualitative case study, the findings need to be interpreted with caution. Our work occurred in a public managed health system, and associated values and motivations may affect generalizability to private providers. Our sample was purposive and focused on clinical leaders and managers and did not capture the perspectives of a broader range of frontline staff. This may be particularly important, considering that our findings highlight the central role of informal networking. We captured perceptions of the importance of informal channels but had limited opportunity to examine the spread and operation of networking among those at the coalface of providing care. In addition, our fieldwork examined knowledge transfer from organizations participating in the GDE program but not organizations outside the program. There is also an overall difficulty in capturing informal knowledge exchanges and a risk that attempts to monitor these will overlook important knowledge transfer processes.

Integration of Findings With the Current Literature

Our findings add to the sparse existing empirical literature exploring learning ecosystems and interorganizational knowledge transfer in digital transformation in health care [24,41,42]. This work highlights the complexity of the health care landscape, involving multiple users and producers of knowledge, driven by various, and at times, conflicting motivations [24,41,42]. Thus, there is no recipe for successful knowledge transfer in innovation ecosystems, and scholars have argued that it is the overall constellation rather than the presence of particular individual factors that determine success [43]. The more informal networking and knowledge transfer becomes, the more organic and self-sustaining it is. Therefore, knowledge transfer and learning cannot be fully centrally planned. It needs to be evolving and shared between central program management and participating organizations.

Understanding how various stakeholder groups acquire and use knowledge and the relative efforts and benefits of using and producing knowledge can help facilitate knowledge transfer [42]. Our work supports the notion that this is a complex and dynamic process characterized by collaboration as well as competition involving various forms of learning [44]. The literature highlights the need for national guidance to stimulate the establishment of a learning ecosystem [42], but there is limited evidence on how this may be achieved and how health systems can organize knowledge transfer more effectively. Here, we provide a starting point for addressing these issues. For example, geographical proximity, communal needs, and technological and cultural similarity can promote interorganizational knowledge transfer, particularly where they enable existing links between organizations and individuals [45,46].

Insights from studies of learning economies in commercial settings are likely to have limited applicability to health, as contexts and underlying processes differ. For example, health system stakeholders are often motivated to exchange knowledge to contribute to public good and patient care, whereas profit is likely to motivate commercial organizations primarily. Nevertheless, there are some areas of convergence that indicate that certain underlying processes are generic. Existing research shows that knowledge transfer across organizations can be achieved using different mechanisms, including databases and codified documents (supporting the importance of the blueprinting process) [47]; workshops and meetings [48]; task forces, visits, and personnel transfers (supporting the importance of informal visits) [44,49,50]; formation of user communities [22]; and formation of alliances (supporting the importance of GDE and fast follower partnerships) [51].

Our results also support existing work highlighting the importance of informal networks and that implementation and optimization experience is difficult to codify and transfer [52-54]. Although different mechanisms vary in effectiveness [55], the most effective way to transfer tacit knowledge is through people interacting and perhaps moving between settings or establishing communities of practice [15,56,57]. Blueprints, as repositories of formal knowledge, can help lower entry costs for neophytes, but they need to be supported by complex informal networking-based approaches that promote different types of learning [58-61].

Moreover, we identified the important emerging role of various intermediaries in knowledge transfer. This intermediary role is often carried out by people, enabled by their location and attitude, rather than through formally managed planned arrangements [19,62,63]. Vendors frequently bring their users together to obtain insights into the context of use of their offerings, which helps them refine and market these [64]. Users can exploit these fora to network and share experience together, thereby securing influence over product enhancement as well as over the strategies adopted by vendors [65]. Professional groupings and some independent organizations geared toward mediating knowledge exchange are also effective forms of intermediaries, as they do not have conflicting interests and therefore do not seek to control members' activities (allowing knowledge to flow freely) [66].

Implications for Policy, Practice, and Research Emerging From This Work

This groundbreaking attempt to create a national digital health learning ecosystem illustrates that formal top-down interventions (such as partnering arrangements, the production of artifacts such as blueprints, funding, and coordination activities) can stimulate the beginnings of a learning ecosystem, but informal relationships (arising from these initiatives or emerging independently) are important for effectiveness and sustainability [67,68]. However, informal networks are difficult to plan, and knowledge transfer and networking cannot be anticipated. Therefore, support should seek to assist informal knowledge markets where formal means have failed. These may include promoting secondments and consultancy to promote knowledge transfer through *social learning*.

Central strategies cannot, however, guarantee that effective informal knowledge transfer will occur; therefore, there is a degree of uncertainty in relation to both intended and unintended outcomes. Policy intervention needs to be evolutionary, establishing ways to help link stakeholders with similar concerns and needs, offering tools to facilitate communication, and encouraging the activity of independent intermediaries [19,56].

Conclusions

Interorganizational knowledge transfer was promoted by formal structures initiated through the GDE program. Informal processes play a key role in knowledge transfer, but they are highly contingent and cannot be readily promoted and sustained by conventional top-down planning structures. National mechanisms to stimulate knowledge sharing, therefore, need to be flexible to align with emerging, changing needs, and need to be sustained through informal networking driven by the mutual benefits of knowledge exchange. Benefits are most immediate and networking most readily sustained where there is strong convergence between group members in their organizational and technological setting and goals, such that the costs of learning are minimized and the benefits of learning are maximized. Recent concerted efforts to deploy digital solutions during the COVID-19 pandemic have reinforced this point [69].

The program laid the foundation for a digital health learning ecosystem. However, interpersonal knowledge transfer (eg, through networking and visits) is labor- and resource-intensive and may be difficult to scale and sustain. Knowledge transfer through circulating documents such as blueprints, although potentially scalable and low cost, is unlikely to be effective by itself. This situation calls for evolving strategic and policy frameworks, shaped by a mixture of top-down and bottom-up inputs, with a trusting relationship between those who facilitate knowledge exchanges and those involved in actively sharing and using that knowledge.

Acknowledgments

The authors gratefully acknowledge the input of the participants and the steering group for this evaluation. This study has drawn on a program of independent research funded by NHS England. The views expressed are those of the authors and not necessarily those of the NHS, NHS England, NHSX, the Department of Health and Social Care, or NHS Digital. This work was also supported by the National Institute for Health Research Imperial Patient Safety Translational Research Centre.

Authors' Contributions

KC and RW conceived the study. KC and RW drafted the manuscript, and all authors commented on the drafts of the manuscript.

Conflicts of Interest

All authors are investigators evaluating the GDE program. AS was a member of the working group that produced Making IT Work and was an assessor in selecting GDE sites. BDF supervises a PhD student partly funded by Cerner, unrelated to this study.

Multimedia Appendix 1
Number of interviews, observations, and documents collected in each case study site.
[DOCX File , 15 KB - jmir_v23i8e23372_app1.docx]

References

1. Buntin MB, Burke MF, Hoaglin MC, Blumenthal D. The benefits of health information technology: a review of the recent literature shows predominantly positive results. Health Aff (Millwood) 2011 Mar;30(3):464-471. [doi: 10.1377/hlthaff.2011.0178] [Medline: 21383365]
2. 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 Med 2011 Jan 18;8(1):e1000387 [FREE Full text] [doi: 10.1371/journal.pmed.1000387] [Medline: 21267058]
3. Five Year Forward View. National Health Service. URL: https://www.england.nhs.uk/wp-content/uploads/2014/10/5yfv-web.pdf [accessed 2020-07-01]
4. The Health Information Technology for Economic and Clinical Health (HITECH) Act. 2009. URL: https://www.healthit.gov/sites/default/files/hitech_act_excerpt_from_arra_with_index.pdf [accessed 2020-07-01]
5. Safe, seamless and secure: evolving health and care to meet the needs of modern Australia. Australia’s National Digital Health Strategy. URL: https://www.digitalhealth.gov.au/sites/default/files/2020-11/Australia%27s%20National%20Digital%20Health%20Strategy%20-%20Safe%2C%20seamless%20and%20secure.pdf [accessed 2020-07-01]
6. Bowden T, Coiera E. Comparing New Zealand's 'Middle Out' health information technology strategy with other OECD nations. Int J Med Inform 2013 May;82(5):87-95. [doi: 10.1016/j.ijmedinf.2012.12.002] [Medline: 23287413]
7. Lau F, Price M, Keshavjee K. From benefits evaluation to clinical adoption: making sense of health information system success in Canada. Healthc Q 2011;14(1):39-45. [doi: 10.12927/hcq.2011.22157] [Medline: 21301238]
8. Halamka JD, Tripathi M. The HITECH era in retrospect. N Engl J Med 2017 Sep 07;377(10):907-909. [doi: 10.1056/nejmp1709851]

9. 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 "early adopter" hospitals. Br Med J 2011 Oct 17;343(oct17 1):d6054 [FREE Full text] [doi: 10.1136/bmj.d6054] [Medline: 22006942]
10. Latest round of technology funding announced. National Health Service. URL: https://www.england.nhs.uk/2015/03/tech-fund-announced/ [accessed 2020-05-06]
11. Cresswell K, Bates D, Sheikh A. Ten key considerations for the successful optimization of large-scale health information technology. J Am Med Inform Assoc 2017 Jan;24(1):182-187 [FREE Full text] [doi: 10.1093/jamia/ocw037] [Medline: 27107441]
12. Cresswell KM, Bates DW, Sheikh A. Ten key considerations for the successful implementation and adoption of large-scale health information technology. J Am Med Inform Assoc 2013 Jun 01;20(e1):9-13 [FREE Full text] [doi: 10.1136/amiajnl-2013-001684] [Medline: 23599226]
13. Secundo G, Toma A, Schiuma G, Passiante G. Knowledge transfer in open innovation. Busi Proc Manag J 2018 Aug 09;25(1):144-163. [doi: 10.1108/bpmj-06-2017-0173]
14. Hamer S. Developing an innovation ecosystem: a framework for accelerating knowledge transfer. J Manag Mark Healthcare 2013 Jul 18;3(4):248-255. [doi: 10.1179/175330310x12736578177607]
15. Lave J, Wenger E. Situated Learning: Legitimate Peripheral Participation. Cambridge Cambridge, United Kingdom: Cambridge University Press; 1991.
16. Lundvall B. Product innovation and user-producer interaction. In: The Learning Economy and the Economics of Hope. London, United Kingdom: Anthem Press; 1985:1-422.
17. New service helps Ontario hospitals with technology. Canadian Healthcare Technology. 2017. URL: https://www.canhealth.com/2017/12/20/new-service-helps-ontario-hospitals-with-technology/ [accessed 2020-05-06]
18. Rogers H, Silvester K, Copeland J. NHS Modernisation Agency's way to improve health care. Br Med J 2004 Feb 21;328(7437):463 [FREE Full text] [doi: 10.1136/bmj.328.7437.463] [Medline: 14976116]
19. Brady T, Tierney M, Williams R. The commodification of industry applications software. Ind Corp Change 1992;1(3):489-514. [doi: 10.1093/icc/1.3.489]
20. Koch C. ERP Software Packages: Between mass-production communities and intra-organizational political processes. In: Technological Change and Organizational Action. Oxfordshire, United Kingdom: Routledge; 2003.
21. Mozaffar H. Entangled biographies: a multi-biographical approach in study of user communities around information infrastructures. In: Proceedings of the 17th European Conference on Research Methodology for Business and Management Studies. 2018 Presented at: ECRM 2018 : 17th European Conference on Research Methodology for Business and Management Studies; Jul 12-13, 2018; Rome, Italy p. 287.
22. Mozaffar H. User communities as multi-functional spaces: innovation, collective voice, demand articulation, peer informing and professional identity (and more). In: The New Production of Users. Oxfordshire, United Kingdom: Routledge; 2016:219-246.
23. Stewart J, Hyysalo S. Intermediaries, users and social learning in technological innovation. Int J Innov Mgt 2011 Nov 20;12(03):295-325. [doi: 10.1142/s1363919608002035]
24. Bougrain F, Haudeville B. Innovation, collaboration and SMEs internal research capacities. Res Policy 2002 Jul;31(5):735-747. [doi: 10.1016/s0048-7333(01)00144-5]
25. von Hippel E. Economics of product development by users: the impact of “sticky” local information. Manag Sci 1998 May;44(5):629-644. [doi: 10.1287/mnsc.44.5.629]
26. von Hippel E. Democratizing innovation: the evolving phenomenon of user innovation. Int J Innov Sci 2009 Mar;1(1):29-40. [doi: 10.1260/175722209787951224]
27. Johnson M, Mozaffar H, Campagnolo GM, Hyysalo S, Pollock N, Williams R. The managed prosumer: evolving knowledge strategies in the design of information infrastructures. Inf Commun Soc 2013 Aug 25;17(7):795-813. [doi: 10.1080/1369118x.2013.830635]
28. Global Digital Exemplars. National Health Service. URL: https://www.england.nhs.uk/digitaltechnology/connecteddigitalsystems/exemplars/ [accessed 2020-05-06]
29. Global Digital Exemplar Programme Evaluation. University of Edinburgh. URL: https://www.ed.ac.uk/usher/digital-exemplars [accessed 2020-05-06]
30. Making IT Work: Harnessing the power of health information technology to improve care in England. Report of the National Advisory Group on Health Information Technology in England. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/550866/Wachter_Review_Accessible.pdf [accessed 2020-06-05]
31. NHS Digital Academy. National Health Service. URL: https://www.england.nhs.uk/digitaltechnology/nhs-digital-academy/ [accessed 2020-05-06]
32. Cresswell K, Krasuska M, Sheikh A, Nguyen H, Hinder S, Franklin B, et al. Inter-organisational knowledge sharing to establish digital health learning ecosystems: qualitative evaluation of a national digital health transformation programme in England. J Med Internet Res 2021:23372. [doi: 10.2196/preprints.23372]

33. Williams R, Sheikh A, Franklin BD, Krasuska M, Nguyen HT, Hinder S, et al. Using Blueprints to promote interorganizational knowledge transfer in digital health initiatives-a qualitative exploration of a national change program in English hospitals. J Am Med Inform Assoc 2021 Mar 11:A. [doi: 10.1093/jamia/ocab020] [Medline: 33706378]
34. Crowe S, Cresswell K, Robertson A, Huby G, Avery A, Sheikh A. The case study approach. BMC Med Res Methodol 2011 Jun 27;11(1):100 [FREE Full text] [doi: 10.1186/1471-2288-11-100] [Medline: 21707982]
35. Cresswell K, Sheikh A, Franklin BD, Krasuska M, Nguyen H, Hinder S, et al. Formative independent evaluation of a digital change programme in the English National Health Service: study protocol for a longitudinal qualitative study. BMJ Open 2020 Oct 08;10(10):e041275 [FREE Full text] [doi: 10.1136/bmjopen-2020-041275] [Medline: 33033100]
36. 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 Jun 10;22(6):e15068 [FREE Full text] [doi: 10.2196/15068] [Medline: 32519968]
37. Patton M. Qualitative Research & Evaluation Methods: Integrating Theory and Practice. Thousand Oaks, California, United States: Sage Publications Inc; 2014:1-832.
38. 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 16;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
39. Global Digital Exemplar Blueprints. National Health Service. URL: https://www.england.nhs.uk/digitaltechnology/connecteddigitalsystems/exemplars/gde-blueprints/ [accessed 2020-05-06]
40. Digital Health Networks is the largest community of NHS IT leaders. Digital Health. URL: https://www.digitalhealth.net/join-the-digital-health-networks/ [accessed 2020-05-06]
41. Stephanie L, Sharma RS. Digital health eco-systems: an epochal review of practice-oriented research. Int J Inf Manag 2020 Aug;53:102032. [doi: 10.1016/j.ijinfomgt.2019.10.017]
42. Laihonen H. A managerial view of the knowledge flows of a health-care system. Knowl Manag Res Pract 2017 Dec 19;13(4):475-485. [doi: 10.1057/kmrp.2014.3]
43. Bacon E, Williams MD, Davies GH. Recipes for success: conditions for knowledge transfer across open innovation ecosystems. Int J Inf Manag 2019 Dec;49:377-387. [doi: 10.1016/j.ijinfomgt.2019.07.012]
44. Inkpen AC, Dinur A. Knowledge management processes and international joint ventures. Organ Sci 1998 Aug;9(4):454-468. [doi: 10.1287/orsc.9.4.454]
45. Cummings JL, Teng B. J Eng Technol Manag 2003 Jun;20(1-2):39-68. [doi: 10.1016/s0923-4748(03)00004-3]
46. Almirall E, Lee M, Majchrzak A. Open innovation requires integrated competition-community ecosystems: lessons learned from civic open innovation. Bus Horiz 2014 May;57(3):391-400. [doi: 10.1016/j.bushor.2013.12.009]
47. Haas MR, Hansen MT. When using knowledge can hurt performance: the value of organizational capabilities in a management consulting company. Strat Mgmt J 2004 Jan;26(1):1-24. [doi: 10.1002/smj.429]
48. Nonaka I, Byosiere P, Borucki CC, Konno N. Organizational knowledge creation theory: a first comprehensive test. Int Busi Rev 1994 Dec;3(4):337-351. [doi: 10.1016/0969-5931(94)90027-2]
49. Almeida P, Kogut B. Localization of knowledge and the mobility of engineers in regional networks. Manag Sci 1999 Jul;45(7):905-917. [doi: 10.1287/mnsc.45.7.905]
50. Gupta AK, Govindarajan V. Knowledge flows within multinational corporations. Strat Mgmt J 2000 Apr;21(4):473-496. [doi: 10.1002/(sici)1097-0266(200004)21:4<473::aid-smj84>3.0.co;2-i]
51. Larsson R, Bengtsson L, Henriksson K, Sparks J. The interorganizational learning dilemma: collective knowledge development in strategic alliances. Organ Sci 1998 Jun;9(3):285-305. [doi: 10.1287/orsc.9.3.285]
52. Nonaka I, von Krogh G. Perspective—tacit knowledge and knowledge conversion: controversy and advancement in organizational knowledge creation theory. Organ Sci 2009 Jun;20(3):635-652. [doi: 10.1287/orsc.1080.0412]
53. Battistella C, De Toni AF, Pillon R. Inter-organisational technology/knowledge transfer: a framework from critical literature review. J Technol Transf 2015 Jun 10;41(5):1195-1234. [doi: 10.1007/s10961-015-9418-7]
54. Fleck J. Contingent knowledge and technology development. Technol Anal Strateg Manag 1997 Jan;9(4):383-398. [doi: 10.1080/09537329708524293]
55. Rosenkopf L, Almeida P. Overcoming local search through alliances and mobility. Manag Sci 2003 Jun;49(6):751-766. [doi: 10.1287/mnsc.49.6.751.16026]
56. Sørensen K. Learning technology, constructing culture. Socio-technical change as social learning. Soc Learn Multimed 1996;18:96. [doi: 10.13140/RG.2.2.30403.71206]
57. Hellström A, Lifvergren S, Gustavsson S, Gremyr I. Adopting a management innovation in a professional organization: the case of improvement knowledge in healthcare. Busi Proc Manag J 2015;21(5):1186-1203. [doi: 10.1108/BPMJ-05-2014-0041]
58. Lundvall B, Johnson B, Andersen ES, Dalum B. National systems of production, innovation and competence building. Res Policy 2002 Feb;31(2):213-231. [doi: 10.1016/s0048-7333(01)00137-8]
59. Almeida P, Song J, Grant RM. Are firms superior to alliances and markets? An empirical test of cross-border knowledge building. Organ Sci 2002 Apr;13(2):147-161. [doi: 10.1287/orsc.13.2.147.534]

60. Szulanski G, Ringov D, Jensen RJ. Overcoming stickiness: how the timing of knowledge transfer methods affects transfer difficulty. Organ Sci 2016 Mar;27(2):304-322. [doi: 10.1287/orsc.2016.1049]
61. Ritala P, Agouridas V, Assimakopoulos D, Gies O. Value creation and capture mechanisms in innovation ecosystems: a comparative case study. Int J Technol Manag 2013;63(3/4):244. [doi: 10.1504/ijtm.2013.056900]
62. van Lente H, Hekkert M, Smits R, van Waveren B. Roles of systemic intermediaries in transition processes. Int J Innov Mgt 2011 Nov 20;07(03):247-279. [doi: 10.1142/s1363919603000817]
63. Kuenne C, Agarwal R. Online innovation intermediaries as a critical bridge between patients and healthcare organizations. Wirtschaftsinformatik Proceedings 2015. 2015. URL: https://aisel.aisnet.org/wi2015/85/ [accessed 2021-07-05]
64. Lager T, Tano K, Anastasijevic N. Open innovation and open production: a case of a technology supplier/user collaboration in the process industries. Int J Innov Mgt 2015 Apr 17;19(02):1550022. [doi: 10.1142/s136391961550022x]
65. Panteli N, Rapti A, Scholarios D. 'If He Just Knew Who We Were': Microworkers' emerging bonds of attachment in a fragmented employment relationship. Work Employ Soc 2020 Jan 27;34(3):476-494. [doi: 10.1177/0950017019897872]
66. Kimble C, Grenier C, Goglio-Primard K. Innovation and knowledge sharing across professional boundaries: political interplay between boundary objects and brokers. Int J Inf Manag 2010 Oct;30(5):437-444. [doi: 10.1016/j.ijinfomgt.2010.02.002]
67. Boland RJ, Lyytinen K, Yoo Y. Wakes of innovation in project networks: the case of digital 3-d representations in architecture, engineering, and construction. Organ Sci 2007 Aug;18(4):631-647. [doi: 10.1287/orsc.1070.0304]
68. Conklin J, Stolee P, Luesby D, Sharratt MT, Chambers LW. Enhancing service delivery capacity through knowledge exchange: the seniors health research transfer network. Healthc Manage Forum 2016 Nov 15;20(4):20-26. [doi: 10.1016/s0840-4704(10)60087-7]
69. Lenert L, McSwain B. Balancing health privacy, health information exchange, and research in the context of the COVID-19 pandemic. J Am Med Inform Assoc 2020 Jun 01;27(6):963-966 [FREE Full text] [doi: 10.1093/jamia/ocaa039] [Medline: 32232432]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Studies
GDE: Global Digital Exemplar
HIT: Health Information Technology
NHS: National Health Service

Edited by R Kukafka; submitted 10.08.20; peer-reviewed by A Coe, S Gajovic, P Nong; comments to author 18.11.20; revised version received 01.12.20; accepted 30.04.21; published 19.08.21.

Please cite as:
Cresswell K, Sheikh A, Franklin BD, Krasuska M, The Nguyen H, Hinder S, Lane W, Mozaffar H, Mason K, Eason S, Potts H, Williams R
Interorganizational Knowledge Sharing to Establish Digital Health Learning Ecosystems: Qualitative Evaluation of a National Digital Health Transformation Program in England
J Med Internet Res 2021;23(8):e23372
URL: https://www.jmir.org/2021/8/e23372
doi:10.2196/23372
PMID:34420927

©Kathrin Cresswell, Aziz Sheikh, Bryony Dean Franklin, Marta Krasuska, Hung The Nguyen, Susan Hinder, Wendy Lane, Hajar Mozaffar, Kathy Mason, Sally Eason, Henry Potts, Robin Williams. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Patterns of eHealth Website User Engagement Based on Cross-site Clickstream Data: Correlational Study

Jia Li[1], PhD; Kanghui Yu[1], BA; Xinyu Bao[1], MA; Xuan Liu[1], PhD; Junping Yao[2], PhD

[1]School of Business, East China University of Science and Technology, Shanghai, China
[2]Xi'an Research Institute of High Technology, Xi'an, China

Corresponding Author:
Junping Yao, PhD
Xi'an Research Institute of High Technology
No.2 Tongxin Road
Xi'an, 710025
China
Phone: 86 18702963951
Email: junpingy200225@163.com

Abstract

Background: User engagement is a key performance variable for eHealth websites. However, most existing studies on user engagement either focus on a single website or depend on survey data. To date, we still lack an overview of user engagement on multiple eHealth websites derived from objective data. Therefore, it is relevant to provide a holistic view of user engagement on multiple eHealth websites based on cross-site clickstream data.

Objective: This study aims to describe the patterns of user engagement on eHealth websites and investigate how platforms, channels, sex, and income influence user engagement on eHealth websites.

Methods: The data used in this study were the clickstream data of 1095 mobile users, which were obtained from a large telecom company in Shanghai, China. The observation period covered 8 months (January 2017 to August 2017). Descriptive statistics, two-tailed *t* tests, and an analysis of variance were used for data analysis.

Results: The medical category accounted for most of the market share of eHealth website visits (134,009/184,826, 72.51%), followed by the lifestyle category (46,870/184,826, 25.36%). The e-pharmacy category had the smallest market share, accounting for only 2.14% (3947/184,826) of the total visits. eHealth websites were characterized by very low visit penetration and relatively high user penetration. The distribution of engagement intensity followed a power law distribution. Visits to eHealth websites were highly concentrated. User engagement was generally high on weekdays but low on weekends. Furthermore, user engagement gradually increased from morning to noon. After noon, user engagement declined until it reached its lowest level at midnight. Lifestyle websites, followed by medical websites, had the highest customer loyalty. e-Pharmacy websites had the lowest customer loyalty. Popular eHealth websites, such as medical websites, can effectively provide referral traffic for lifestyle and e-pharmacy websites. However, the opposite is also true. Android users were more engaged in eHealth websites than iOS users. The engagement volume of app users was 4.85 times that of browser users, and the engagement intensity of app users was 4.22 times that of browser users. Male users had a higher engagement intensity than female users. Income negatively moderated the influence that platforms (Android vs iOS) had on user engagement. Low-income Android users were the most engaged in eHealth websites. Conversely, low-income iOS users were the least engaged in eHealth websites.

Conclusions: Clickstream data provide a new way to derive an overview of user engagement patterns on eHealth websites and investigate the influence that various factors (eg, platform, channel, sex, and income) have on engagement behavior. Compared with self-reported data from a questionnaire, cross-site clickstream data are more objective, accurate, and appropriate for pattern discovery. Many user engagement patterns and findings regarding the influential factors revealed by cross-site clickstream data have not been previously reported.

(J Med Internet Res 2021;23(8):e29299) doi:10.2196/29299

KEYWORDS

engagement; clickstream data; cross-site visit; platform; channel; mobile phone

Introduction

Background

Providing and delivering web-based services is a major trend in the digital transformation of health services [1]. Currently, increasingly more users obtain health information, consult a physician, purchase drugs, or self-manage wellness from smartphones. As a result, understanding user behavior on smartphones is becoming increasingly important. Owing to differences in screen size and mobility, users behave differently on smartphones than on desktop computers [2]. Therefore, it is relevant to investigate eHealth website usage behaviors such as user engagement from smartphones.

User engagement is a key variable for eHealth websites [3]. The sizable demand for web-based health services has led to a large number of eHealth websites, which makes competition extremely fierce. According to the IQVIA Institute, more than 318,000 health apps are now available on top app stores worldwide, with more than 200 health apps being added each day [4]. As a result, it is increasingly difficult to obtain sufficient engagement for users on specific health websites. In addition, health care is a relatively low-frequency need compared with social networking, news reading, or web-based shopping. Users visit health websites or apps only when they have health concerns. All these factors make achieving sufficient user engagement on eHealth websites a difficult task.

Although many previous studies have investigated engagement patterns [5-9] and engagement interventions [3,10-15] on eHealth websites, most of them only focus on a single website. As a result, the research findings can only be applied to the corresponding categories of the eHealth website. However, the patterns of user engagement (eg, market share, penetration, intensity, variety, time trends, loyalty, and cross-site visits) on all eHealth websites and the factors that influence user engagement on all eHealth websites need to be examined. In addition, the links among the different categories of eHealth websites are largely unknown. For example, questions such as how users visit multiple eHealth websites simultaneously or how one type of eHealth website can provide referral traffic for other types of eHealth websites have not been answered by previous studies. By solving these questions, we can better understand user engagement behavior from a holistic view and keep users more engaged in different types of eHealth websites.

Literature Review

User engagement on eHealth websites has received considerable attention in recent years. A review of the literature suggests two main research streams investigating engagement patterns and engagement interventions. The first research stream is descriptive in nature. The areas investigated include diabetes management [5,6], mental health management [7], pain management [8], and health information dissemination [9]. For diabetes management, Böhm et al [5] found that although more women used the app, they engaged significantly less with it. Older people and users who were recently diagnosed tended to use apps more actively. Glasgow et al [6] investigated engagement patterns on diabetes self-management websites. They found that participants visited the website fairly often and used all of the theoretically important sections, but engagement decreased over 4 months. For mental health management, Baumel et al [7] found that daily minutes of use were significantly higher for mindfulness, meditation and peer support apps than for apps incorporating other techniques (tracker, breathing exercise, and psychoeducation). The median 15-day and 30-day retention rates of the app were 3.9% and 3.3%, respectively, indicating that only a small portion of users actually used the apps for a long period. For pain management, Rahman et al [8] found that although most users of the app reported being female, male users were more likely to be highly engaged in the app. Users in the most engaged clusters self-reported a higher number of pain conditions, a higher number of current medications, and a higher incidence of opioid usage. For health information dissemination, Zhang et al [9] investigated the user engagement of health information disseminated by Chinese provincial centers for disease control and prevention on WeChat. They found that the median number of reads was 551.5 and the median number of likes was 10. Article content, article type, communication skills, number of marketing elements, and article length were associated with the reading and liking levels. However, title type was only associated with liking level.

The second research stream focuses on designing interventions to improve user engagement. System design [3,10], social support [11,12], gamification [13,14], and channels [15] are the most frequently investigated interventions to promote user engagement. For the system design, Baumel and Kane [10] found that therapeutic persuasiveness, therapeutic alliance, visual design, and content predict an increase in user engagement with eHealth interventions. Wei et al [3] investigated which design features improved user engagement with mobile health interventions. They identified the following seven themes that influenced user engagement: personalization, reinforcement, communication, navigation, credibility, message presentation, and interface esthetics. For social support, Kashian and Jacobson [11] found that optimal social support and tie strength were positively related to engagement. In addition, the more engaged members were, the more positive their health expectations were. Wang et al [12] revealed that the amount and match of received support were positive and significant predictors of new users' continued engagement. For gamification, Edney et al [13] found that the inclusion of gamified features enhanced engagement in an app-based physical activity intervention. Comello et al [14] found that a game-inspired infographic showed the potential to outperform a traditional format for comprehension and decreased cognitive load while not underperforming on engagement (eg, attitudes and emotional tone). With regard to the channel used, Brusk and Bensley [15] compared the impact of mobile versus fixed devices on user engagement key performance indicators. They found that eight user characteristics (lessons completed, race, ethnicity, language, state of residence, pregnancy status, beginning stage of change, and preferred nutrition education method) were significantly related to various key performance indicator differences between mobile and fixed device access. Surprisingly, their results suggest that nonmobile users are more likely to engage.

The review results listed above indicate that extant studies on user engagement in eHealth only focus on a single website. A higher level of analysis that provides a complete picture of how users engage in different types of eHealth websites is still lacking. Although the meta-analysis allows multiple websites to be considered together, existing review studies on this topic still focus on a single category [3]. The link between the different categories of eHealth websites is missing. To bridge this gap, we investigate user engagement behavior on all eHealth websites based on cross-site clickstream data.

Research Questions

To bridge this research gap, we provide an analysis of user engagement on all eHealth sites with cross-site clickstream data in this study. Following the two research streams on user engagement [5-15], we focus on both user engagement patterns and engagement interventions.

First, we are interested in investigating user engagement patterns on all eHealth websites. More specifically, we will provide a framework for understanding the engagement patterns on eHealth websites. The framework includes the taxonomy of eHealth websites, market share, penetration, engagement intensity, engagement variety, day and hour trends, customer loyalty, and cross-site engagement. The taxonomy of eHealth websites is necessary because there are too many individual eHealth websites that cannot be covered in a single study. In addition, working on specific websites makes it difficult to reach a conclusion with general significance. Market share and penetration are included because they can jointly describe the market status quo and potential for that type of eHealth website (eg, a small market share with a high penetration usually means a great potential). Intensity, variety, time trend, loyalty, and cross-site behavior are included because they describe different aspects of user engagement. Therefore, the first research question (RQ) is as follows: What are the overall patterns of user engagement on multiple eHealth websites on smartphones (RQ1)?

Second, we are interested in identifying the factors that may influence user engagement on all eHealth websites. Following the Person, Environment, and Technology framework [16], user behavior in information systems can be well explained by personal, environmental, and technological factors. For personal factors, we focused on sex because male and female users exhibit sizable differences in their web behavior [17]. For environmental factors, we focused on income because the literature has suggested that behavior differences exist between high-income and low-income users [17]. For technological factors, we focused on the platform (operating system of the mobile phone) and channel (mobile browsers or mobile apps). Platform was included because there are many differences between iOS and Android communities, such as the number of eHealth apps and the percentage of free apps, which may lead to different engagement behaviors on eHealth websites [18]. Channel was also included because an app channel provides better experience than a browser channel, and such an advantage may lead to more intensive user engagement. To sum up, we investigate how the platform, channel, sex, and income influence user engagement on eHealth websites. Therefore, the second RQ is as follows: How do factors such as the platform, channel, sex, and income influence user engagement on multiple eHealth websites on smartphones (RQ2)?

This study has several practical implications. First, our clickstream data analysis indicates that the visit penetration for eHealth websites is very low, and users usually concentrate only on one or two websites. However, eHealth websites are also characterized by relatively high user penetration. Therefore, eHealth websites should have great market potential. One possible way to increase user engagement on more eHealth websites is to provide cross-site recommendations. Medical websites are ideal sources for effectively providing referral traffic for lifestyle and e-pharmacy websites. However, managers must be cautious that the opposite may not be true. Understanding the asymmetric nature of cross-site browsing can help managers improve the effects of cross-site recommendations.

Second, the findings of this research show that Android users are more engaged in eHealth websites than iOS users, partly because more health apps are available on the Android platform, and the Android platform has a higher percentage of free apps than the iOS platform. Therefore, managers of the iOS platform should encourage developers to develop more health apps (especially free apps or apps with in-app purchase features) in the future.

Third, the results of this study suggest that app users are, on average, 4.5 times more engaged than browser users. Therefore, all eHealth websites should provide apps for both Android and iOS platforms. The managers of eHealth websites should also encourage users to download their apps and urge users to access their websites from apps instead of browsers.

Methods

The data used in this study are the access log data of 1095 4G users from a large telecom company in Shanghai, China. The observation period was 8 months (January 2017-August 2017). After removing confidential information (eg, telephone numbers), we obtained users' internet access records on smartphones. Each access record contains the encrypted user ID, access time, mobile platform (mobile operating system), and URL visited. User demographic information such as encrypted user ID, sex, age, and monthly expenditures on mobile phones was also included in the data set.

The eHealth websites investigated in this study can be classified into the following three categories: medical, lifestyle, and e-pharmacy [19]. Medical websites provide medical information on specific diseases or treatments. Lifestyle websites provide health information on fitness, weight loss, health management, or beauty care. e-Pharmacy websites provide web-based pharmacy services. For each category, we included the top 79.99% (184,826/231,033) most visited websites as the targets in this study, as per the 80-20 rule (also known as the Pareto principle); 20% of websites accounted for 80% of visits. This allowed us to investigate a small number of websites and obtain good coverage of visits to all eHealth websites. By following this approach, the eHealth websites investigated in this study were identified; they are listed in Table 1. The indicators (ie, visits and visitors) in Table 1 were calculated based on all 373 users who visited the websites listed in Table 1 during the 8-month observation period (January 2017 to August 2017).

Table 1. The eHealth websites investigated in this study (in China; n=373).

Category and website	Domain name	Visits, n (%)	Visitors, n (%)
Medical			
Good Doctor	haodf.com	44,202 (23.9)	56 (15.1)
WeDoctor	guahao.com	37,734 (20.4)	38 (10.2)
39 Health Net	39.net	23,007 (12.5)	42 (11.4)
Ask Doctor Quickly	120ask.com	15,613 (8.5)	70 (18.9)
Seeking Medical Advice	xywy.com	11,769 (6.4)	69 (18.6)
Chunyu Doctor	chunyuyisheng.com	1684 (0.9)	9 (2.3)
Lifestyle			
Mint Health	boohee.com	41,462 (22.4)	8 (2.1)
Health Preserving	cndzys.com	4307 (2.3)	17 (4.6)
So-Young	soyoung.com	1101 (0.6)	34 (9)
e-Pharmacy			
Kang Aiduo Pharmacy	360kad.com	2639 (1.4)	11 (3)
Jianke Pharmacy	jianke.com	1308 (0.7)	17 (4.7)

Results

User Engagement Patterns

The engagement patterns investigated in this study include market share, penetration, engagement intensity, engagement variety, day and hour trends, customer loyalty, and cross-site engagement.

Market Share

Market share is the percentage of the market that a single category controls based on the number of visits. The proportion of medical websites is relatively large and accounts for 72.51% (134,009/184,826) of the total, whereas the proportion of e-pharmacy websites is very small and accounts for only 2.14% (3947/184,826) of the total. The proportion is lifestyle websites is 25.36% (46,870/184,826).

This finding indicates that the greatest demand for eHealth websites is to obtain health knowledge and medical advice such as that on prevention, diagnosis, prognosis, and treatment plans. Lifestyle websites also received considerable market share, suggesting that the idea of health management is currently pervasive in China. However, the proportion of visits to e-pharmacy websites was relatively small. A possible reason for this is that the purchase of drugs is a low-frequency demand. Another possible reason is that e-pharmacies are not yet included in the scope of medical insurance in most areas of China. Lack of trust in e-pharmacy websites is also a reason.

eHealth Behavior Penetration

eHealth behavior penetration measures how user behaviors on eHealth websites compare with those of all web behaviors. We focus on two types of user behaviors (Table 2): visit penetration (the percentage of visits to eHealth websites with respect to the visits to all websites) and user penetration (the percentage of users who have ever visited an eHealth website). The results in Table 2 show that the visit penetration of eHealth behavior is quite low, suggesting that eHealth websites correspond to very low-frequency demand compared with all web-based demands (eg, social networking, reading news, and web-based shopping). Users will access eHealth websites only when they have health concerns.

The results in Table 2 also suggest that the user penetration of eHealth behavior is relatively high (142/373, 38.1%). Overall, 33.5% (124/373) of the users had visited a medical website, 12.7% (47/373) had visited a lifestyle website, and 6.1% (23/373) had visited an e-pharmacy website. Considering that we only include 11 top-ranked eHealth websites in this study, the actual user penetration rates will be higher than the estimation reported in Table 2. Therefore, eHealth is an application with low traffic but high user penetration, which is closely related to everyone and has great market potential.

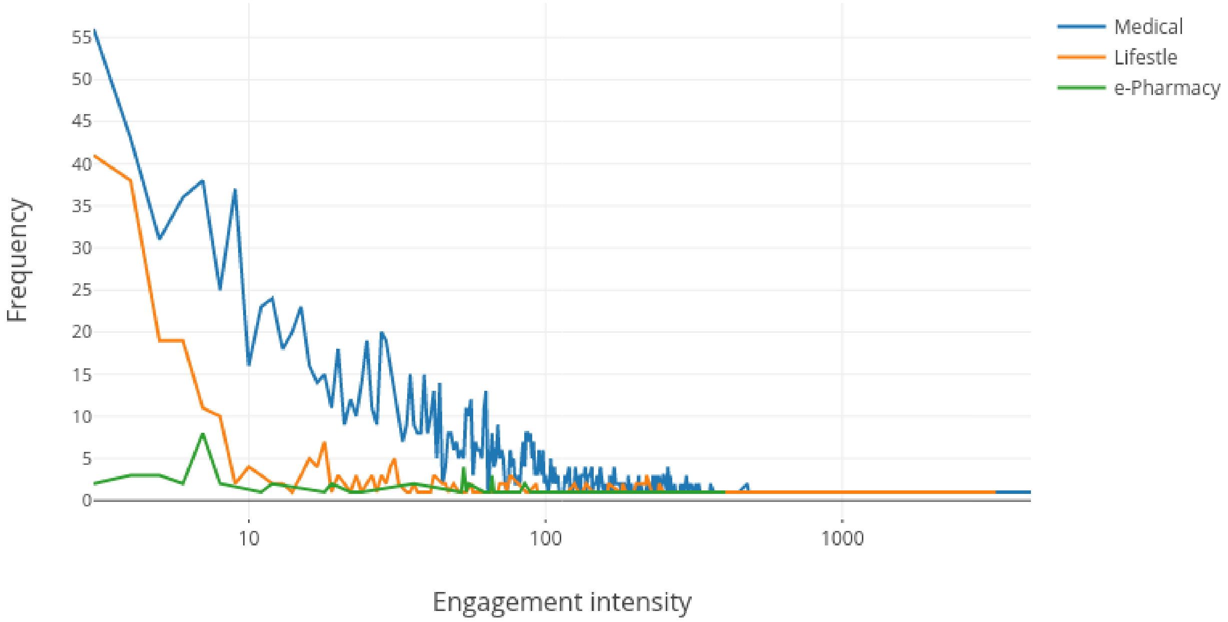
Table 2. eHealth behavior penetration among three categories (n=373).

Category	Visit penetration, n (%)	User penetration, n (%)
Medical	134,009 (0.082)	124 (33.5)
Lifestyle	46,870 (0.029)	47 (12.7)
e-Pharmacy	3947 (0.002)	23 (6.1)

Engagement Intensity

Engagement intensity is the number of visits to eHealth websites per session. In this study, we defined the length of a session as a day. Therefore, we measured engagement intensity as the number of visits to eHealth websites within a day. The engagement intensity patterns according to category are shown in Figure 1. As shown in Figure 1, the frequency decreased exponentially as the engagement intensity increased. The average engagement intensity is 105, indicating that a typical user interacts with these websites 105 times per day to fulfill their needs.

Figure 1. Visit intensity per day of the three categories of eHealth websites.



The results of the Kolmogorov-Smirnov test (D=0.025; *P*=.93) suggest that the distribution of engagement intensity follows a power law distribution [20]. That is, a large number of eHealth needs involve only a small number of visits, whereas a very small number of complex eHealth needs must be realized through a large number of visits. The mechanism of the power law distribution is the lack of natural growth constraints. The number of Facebook fans, the distribution of wealth in an unfair society, or the number of hits on web pages are all examples of data that follow a power law distribution. Visits to eHealth websites also lack constraints. Users can visit the website as many times as they want. However, a majority of health needs are simple. In most cases, users only need to visit eHealth websites 5-10 times to satisfy their needs.

Engagement Variety

Engagement variety measures the extent to which users visit different types of eHealth websites. In this study, engagement variety was measured by the number of distinct eHealth websites over 3 months (Table 3). The results in Table 3 suggest that most users (238/373, 63.8%) visited only 1 eHealth website within 3 months. On average, each user visits 1.5 of the eHealth websites in 3 months. Fewer than 40% (135/373, 36.2%) of users visit multiple eHealth websites. Among these users, 90% (121/135, 89.6%) visit only two to three websites, and very few users visit four or more websites.

Table 3. The distribution of engagement variety (n=373).

Number of websites accessed	Visitors, n (%)
1	238 (63.8)
2	76 (20.4)
3	45 (12.1)
4	7 (1.9)
5	5 (1.3)
6	2 (0.5)
7	0 (0)

This finding suggests that eHealth websites are highly isolated. Users have great inertia and pay attention to only one or two websites. For example, low engagement variety may be attributed to the fact that increasingly more eHealth websites provide one-stop services where users can meet almost all their health needs on one site. The low visit variety also suggests that the links among eHealth websites are insufficient. As a result, users from one website may not be aware of other websites for quite a long time.

Day of the Week Trends

We were interested in user engagement patterns at the week and day levels. For both the week and day levels, we observed the trends of the three key engagement variables (ie, engagement volume, user volume, and engagement intensity) over time. The engagement volume was measured by the number of visits. The user volume was measured by the number of unique users. The engagement intensity was measured by the number of visits per user. All the measures for engagement volume, user volume, and engagement intensity were based on 373 users who visited the websites listed in Table 1 (January 2017-August 2017). The engagement trends at the week level are shown in Figures 2-4.

Figure 2. The fluctuation of the engagement volume in a week.

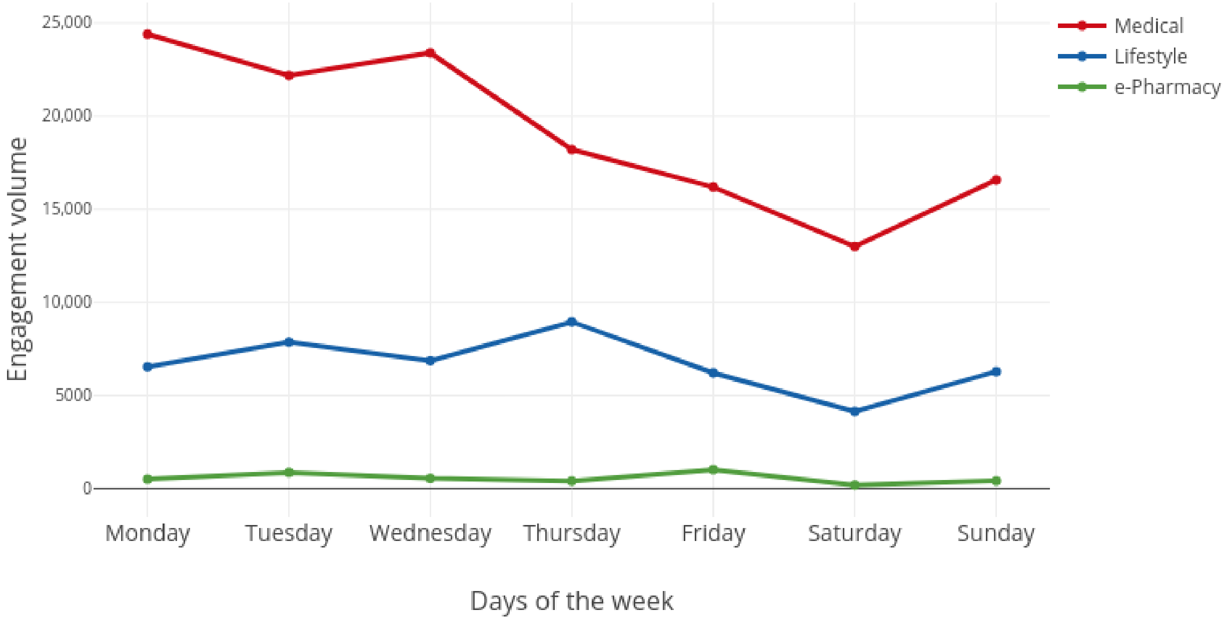


Figure 3. The fluctuation of the user volume in a week.

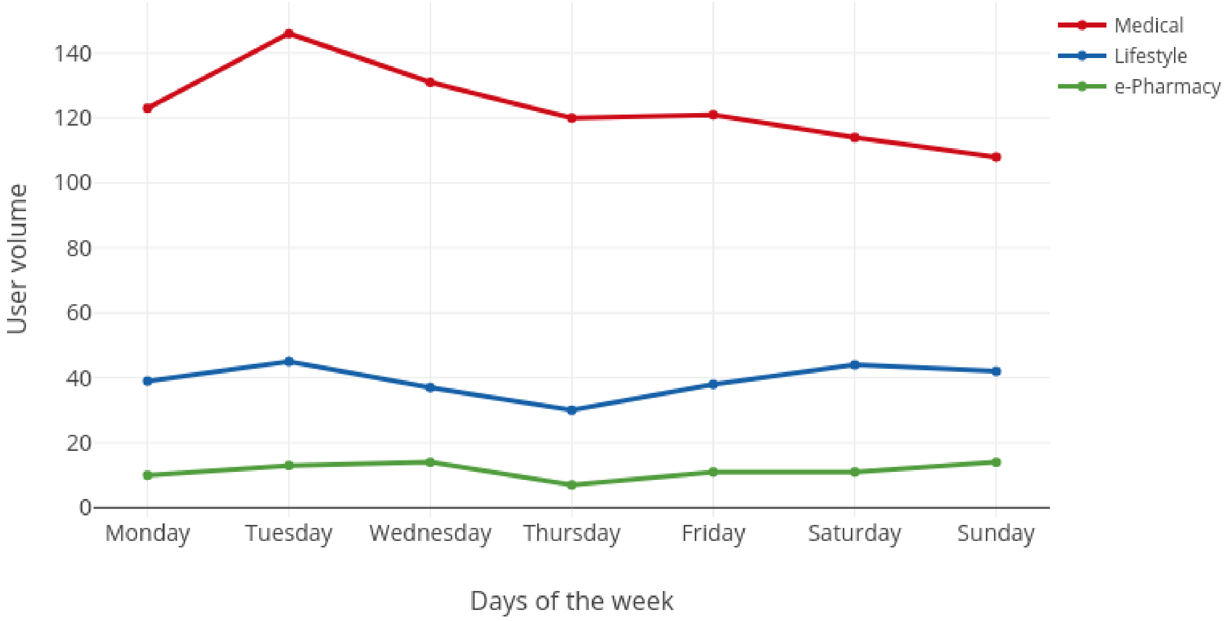
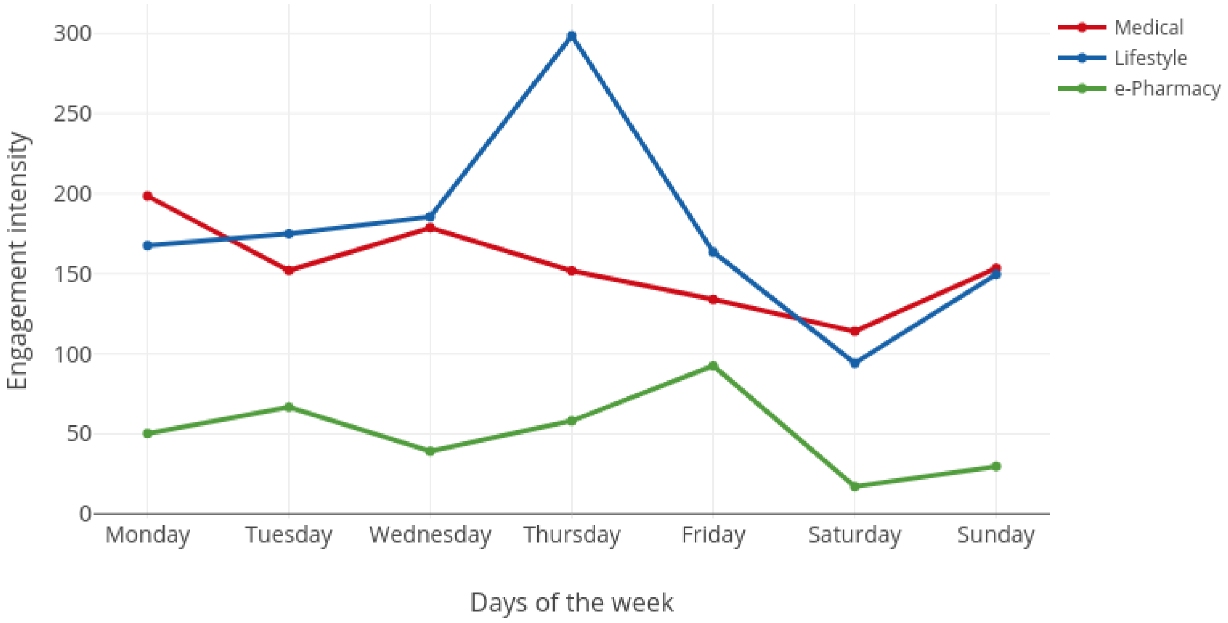


Figure 4. The fluctuation of the engagement intensity in a week.



For medical websites, there was more engagement from Monday to Wednesday, with the highest engagement volume and intensity seen on Monday. From Thursday to Saturday, the engagement volume and intensity decreased gradually until Sunday. The user volume was the highest on Tuesday, but the lowest on Sunday. In addition, the user volume of medical websites fluctuated more than that of the other two categories of websites.

For lifestyle websites, the engagement volume and intensity increased from Sunday to Thursday and then gradually decreased until Saturday. Engagement volume and intensity were the highest on Thursday and lowest on Saturday. However, the user volume on Thursday was the lowest in the week.

For e-pharmacy websites, the engagement volume, user volume, and engagement intensity were all the lowest compared with those of medical and lifestyle websites. The engagement volume and intensity were the highest on Friday but lowest on Saturday.

As shown in Figures 2-4, user engagement is generally high on weekdays but low on weekends. This finding is consistent with previous findings in the social network context that the posting of microblogs is usually more intensive on weekdays than on weekends [21]. One possible explanation is that users are more interested in offline relaxation activities on weekends.

Hour of the Day Trends

The engagement trends at the day level are shown in Figures 5-7. As shown in Figures 5-7, user engagement increased gradually from morning to noon. After noon, user engagement declined until it reaches its lowest level at midnight. In other words, user engagement reached the highest level around noon and the lowest level at midnight. The main reason for this pattern is that users are prone to check their phones during midday lunch hours. In contrast, users engage the least at midnight because they fall asleep at that time.

Figure 5. The fluctuation of the engagement volume in a day.

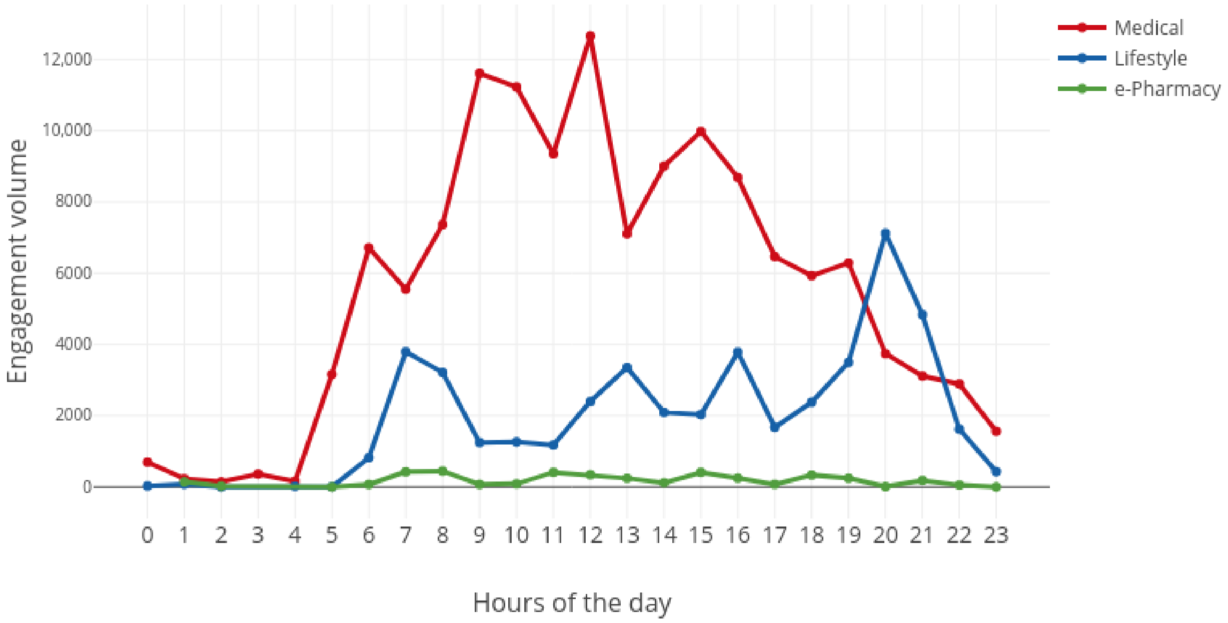


Figure 6. The fluctuation of the user volume in a day.

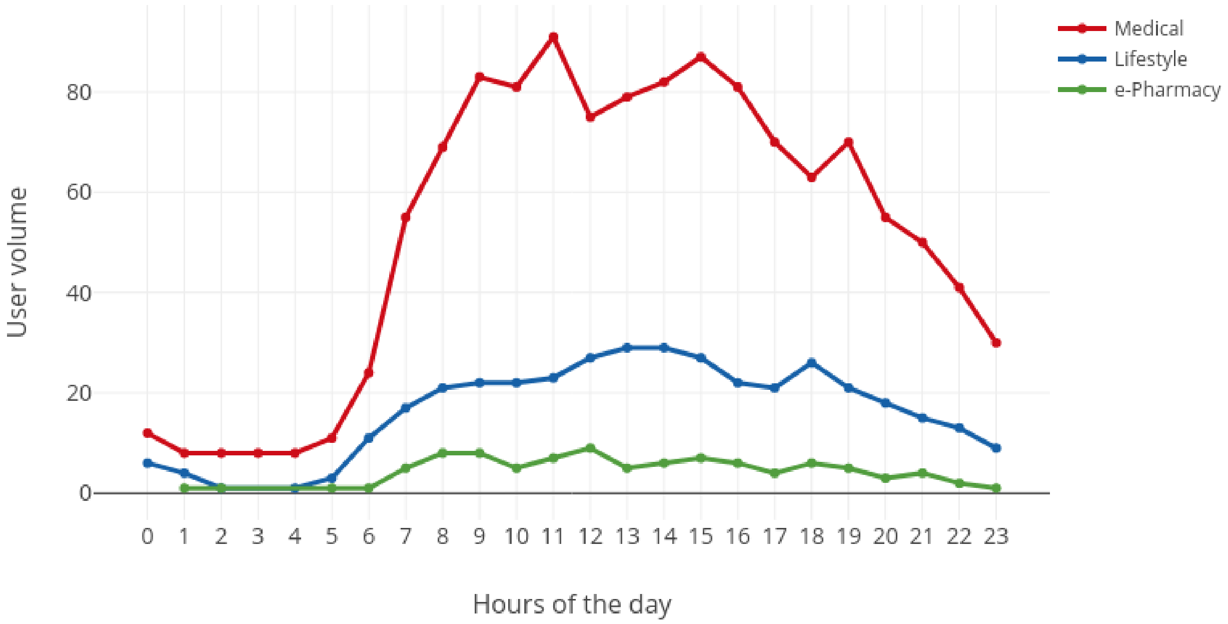
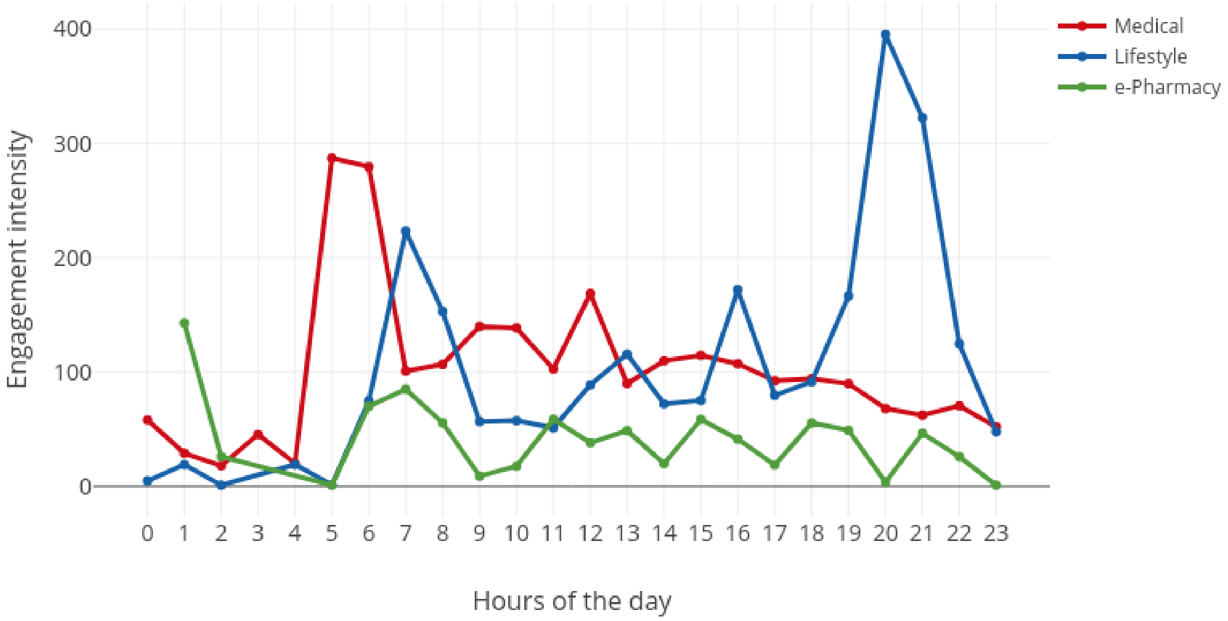


Figure 7. The fluctuation of the engagement intensity in a day.



For medical websites, user engagement peaked at noon. The highest engagement volume appeared at 12 noon, and the highest user volume appeared at 11 AM. However, peak engagement intensity occurred between 5 AM and 6 AM. One possible explanation is that users who encounter health problems at night will search for health information on the web during this period.

For lifestyle websites, the highest engagement was in the evening. For example, peak engagement volume and intensity occurred at 8 PM. This is because the use of lifestyle websites (eg, yoga exercise) usually takes a long time, and the ideal time is right after work. However, the largest number of users were engaged in lifestyle websites at 6 PM. Other peaks in engagement volume and intensity occurred at 7 AM, 1 PM, and 4 PM.

For e-pharmacy websites, user engagement fluctuated throughout the day. One special case is that the engagement intensity reaches its peak at 1 AM. e-Pharmacies are the most intensively used eHealth sites, and they have an engagement intensity that is even higher than those of the remaining two categories (ie, medical and lifestyle). One possible explanation for this phenomenon is that offline drug stores are closed at this time, and e-pharmacies are the only choice.

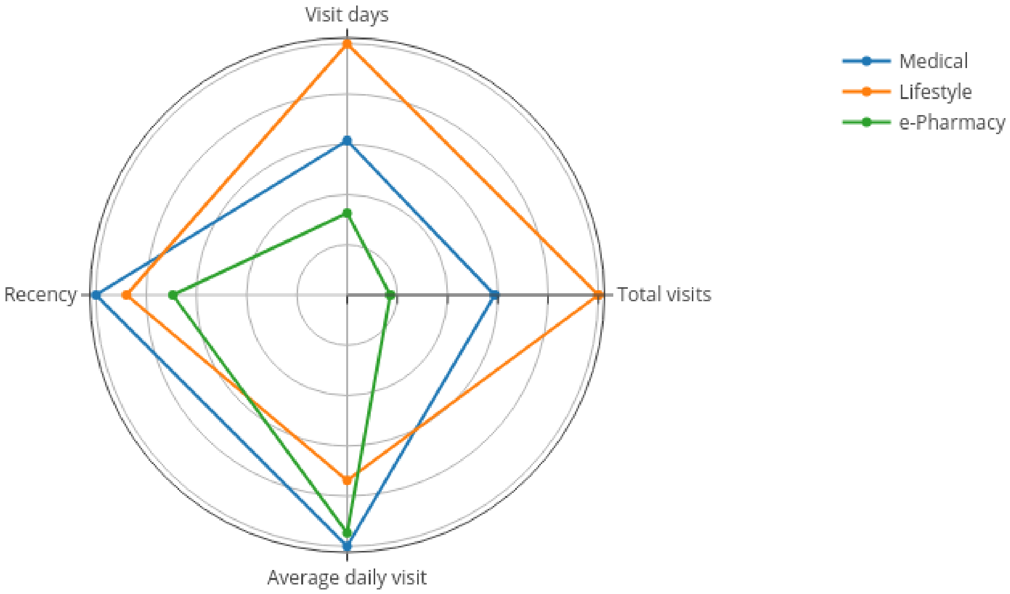
Customer Loyalty

Customer loyalty is a measure of a customer's likelihood of engaging in repeat business with a company or brand. In this study, customer loyalty was measured using the following variables:

1. Total visits: the total number of visits within the observation period.
2. Visit days: the number of days visited within the observation period.
3. Average daily visits: the average number of visits per day.
4. Recency: the number of days since the last visit (in this study, recency was measured based on the difference between the last visit date and the end of the observation period).

A radar chart was used to present customers' loyalty to the three categories of eHealth websites (Figure 8). A radar chart consists of a sequence of equiangular spokes called *radii*, with each spoke representing one of the variables. The data length of a spoke is proportional to the magnitude of the variable for the data point relative to the maximum magnitude of the variable across all data points. A line is drawn connecting the data values for each spoke. This gives the plot a starlike appearance and is the origin of one of the popular names for this plot.

Figure 8. Customer loyalty.



The results in Figure 8 suggest that lifestyle websites, followed by medical websites, have the highest customer loyalty. e-Pharmacy websites have the lowest customer loyalty. This is because the service provided by lifestyle websites requires consistent user engagement over time (eg, fitness and chronic disease self-management). Medical websites also have relatively high customer loyalty because they act as portals to many health services such as health education, web-based consultation, and web-based registration. e-Pharmacies have the lowest customer loyalty because the demand for web-based drug purchases can often be fulfilled by offline pharmacies or hospitals. Many patients visit a web-based pharmacy only when they cannot obtain the drugs offline. In addition, different e-pharmacies are also substitutes for each other because the drugs they sell are standard products.

Cross-site Engagement

A user may visit several eHealth websites simultaneously, a phenomenon known as cross-site visits [22]. In this study, we were interested in two levels of cross-site engagement: user level and visit level.

At the user level, suppose that the number of users who visit lifestyle websites is x and the number of users who also visit medical websites is y. The cross-site engagement of lifestyle websites with medical websites is y/x. Cross-site engagement at the user level is shown in Table 4. The cross-site engagement of users of lifestyle websites to medical websites was 0.66, indicating that 66% (31/47, 66%) of lifestyle users also visited medical websites. Similarly, the cross-site engagement of users of e-pharmacy websites to medical websites was 0.92, and the cross-site engagement of users of e-pharmacy websites to lifestyle websites was 0.46. It should be noted that cross-site browsing was not symmetrical. As shown in Table 4, 66% (31/47, 66%) of lifestyle website users visited medical websites, whereas only 25% (31/124, 25%) of medical website users visited lifestyle websites. The asymmetric nature indicates that medical websites are more popular than lifestyle websites, and lifestyle websites are more popular than e-pharmacy websites.

Table 4. Cross-site engagement on the user level.

Category	Medical, n/N (%)	Lifestyle, n/N (%)	e-Pharmacy, n/N (%)
Medical	124/124 (100)	31/124 (25)	21/124 (17)
Lifestyle	31/47 (66)	47/47 (100)	11/47 (23)
e-Pharmacy	21/23 (92)	11/23 (48)	23/23 (100)

At the visit level, suppose that the number of users who visit the lifestyle websites is w, and the corresponding number of visits is u. On the same day, the number of visits to medical websites by these users is v, and the cross-site engagement of lifestyle websites to medical websites is v/u. Cross-site engagement at the visit level is shown in Table 5. The cross-site engagement of visits to lifestyle websites to medical websites was 0.94, indicating that 94% (44,223/46,870, 94%) of the visits to lifestyle websites were associated with visits to medical websites on the same day. Similarly, the cross-site engagement of visits to e-pharmacy websites to medical websites was 0.98, and the cross-site engagement of visits to e-pharmacy websites to lifestyle websites was 0.71. However, the cross-site engagement of visits was much lower for the opposite case. For example, the cross-site engagement of visits to medical websites to lifestyle websites was 0.33, the cross-site engagement of visits to medical websites to e-pharmacy websites was 0.03, and the cross-site engagement of visits to lifestyle websites to e-pharmacy websites was 0.06. This finding indicates that popular eHealth websites such as medical websites can effectively provide referral traffic for lifestyle and e-pharmacy websites. However, the opposite is not true. Lifestyle and e-pharmacy websites can provide only limited referral traffic for medical websites.

Table 5. Cross-site engagement on the visit level (n=373).

Category	Medical, n/N (%)	Lifestyle, n/N (%)	e-Pharmacy, n/N (%)
Medical	134,009/134,009 (100)	44,223/134,009 (33)	3868/134,009 (3)
Lifestyle	44,223/46,870 (94)	46,870/46,870 (100)	2812/46,870 (6)
e-Pharmacy	3868/3947 (98)	2812/3947 (71)	3947/3947 (100)

Factors Influencing User Engagement

Overview

In this section, we investigate how the platform, channel, sex, and income influence user engagement (ie, engagement volume and engagement intensity) on eHealth websites. More specifically, we first investigate their influence independently and then investigate their interaction effects. Engagement volume is measured by the number of visits, and engagement intensity is measured by the number of visits per session (in this study, a session is defined as 1 day).

Platform

The platform refers to the operating system of the mobile phone used to visit eHealth websites. In this study, we focus on two platforms, iOS and Android, because they possess 97% of the global mobile market share. There are many differences between iOS and Android that may lead to different engagement behaviors on eHealth websites. Android has the greatest global market share at approximately two-thirds and has more app downloads than iOS. Sensor Tower reports that the Google Play Store experienced approximately 75.7 billion first-time app installs worldwide in 2018 [23]. Comparatively, the App Store experienced only 29.6 billion such installs. The Android platform also has more apps than the iOS platform (2.7 million Android apps vs 1.8 million iOS apps) [18]. In addition, Google Play Store has a higher percentage of free apps than the App Store [18].

iOS and Android also have different user groups. Owing to its broad price range and lower entry-level price point, Android has the largest global share in lower-income areas and developing nations [24]. It holds an advantage over iOS in emerging markets such as Asia and Africa. There is also a large gap between the purchasing power of an average iPhone user and that of an Android user. The median iPhone app user earns US $85,000 per year, which is 40% more than the median annual income of Android phone users (US $61,000) [24]. In addition, even though Android users have far more downloads than iOS users, iPhone users spend twice as much as their Android counterparts [24]. Android users also differ from iPhone users in their personalities. According to a recent study, Android users are less extroverted than iPhone users and are perceived to have greater levels of honesty and humility [25].

The results of the two-tailed *t* test comparing engagement volume and engagement intensity between Android and iOS users are shown in Table 6. In addition to the results of the *t* tests, we also report the effective sizes (small, medium, large, very large, and huge) to indicate the magnitudes of the differences. The effect size was first measured using Cohen *d* [26] and then interpreted as small (<0.01), medium (0.01-0.20), large (0.20-0.50), very large (0.50-0.80), or huge (0.80-2.0) according to the values of Cohen *d* [27]. The results in Table 6 show that Android users have a larger engagement volume (Cohen d=0.23; t_{371}=2.26; P=.02, large effect size) and engagement intensity (Cohen d=1.39; t_{371}=32.10; P<.001, large effect size) than iOS users.

Table 6. Comparison of user engagement between platforms[a].

Platform engagement	Value, mean (SD)	*t* test (*df*)	*P* value	Cohen *d*
Engagement volume		2.26 (371)	.02	0.23
Android	3.98 (2.08)			
iOS	3.48 (2.22)			
Engagement intensity		32.10 (371)	<.001	1.39
Android	3.29 (1.90)			
iOS	1.18 (1.00)			

[a] Box-Cox transformation was applied to engagement volume and engagement intensity.

One possible explanation for the difference is that there are more health apps and fewer charges on Android. This makes it easier for Android users to find free health apps to satisfy their needs. In addition, Android users are more introverted and more proficient in information technology [25]. As a result, they are more willing to solve health concerns by visiting eHealth websites. In contrast, iPhone users may be more willing to go offline because of their health concerns.

Channel

The channel refers to the method through which a mobile user interacts with an eHealth website. In this study, we focused on two types of channels: mobile browsers and mobile apps. A browser can be found on any mobile phone, regardless of the operating system. Accessing an eHealth website through a browser is convenient because users do not need to download or install an app before the visit. However, it is essential to remember that network access, quality, and speed are all factors that can affect mobile web experience. Compared with a browser, an app has several advantages. For example, mobile apps offer greater personalization and operational efficiency, along with multiple other exclusive features. A well-designed mobile app can perform actions much quicker than a mobile website. In contrast to websites that generally use web servers, apps usually store their data locally on mobile devices. For this reason, data retrieval is quicker on mobile apps. Apps can further save users’ time by storing their preferences and using them to take proactive actions on their behalf. In addition, mobile apps can access and use built-in device features such as cameras, GPS, and location. Leveraging device capabilities leads to an enhanced, more convenient user experience.

We performed a *t* test to compare the engagement volume and engagement intensity between mobile browser users and mobile app users, and the results are shown in Table 7. The results in Table 7 show that app users have a larger engagement volume (Cohen *d*=1.44; t_{371}=15.51; *P*<.001, large effect size) and engagement intensity (Cohen *d*=1.09; t_{371}=21.51; *P*<.001, huge effect size) than browser users. The engagement volume of app users is 4.85 times that of browser users, and the engagement intensity of app users is 4.22 times that of browser users. Convenient access without remembering website URLs, faster and more fluent user experience, more powerful functions (eg, location-based service, alerts, and Quick Response code scanning), and more customization and personalization are all advantages of apps that may explain such a huge engagement gap.

Table 7. Comparison of user engagement between channels[a].

User engagement and channel	Value, mean (SD)	*t* test (*df*)	*P* value	Cohen *d*
Engagement volume		15.51 (371)	<.001	1.44
App	3.78 (2.13)			
Browser	0.78 (0.76)			
Engagement intensity		21.51 (371)	<.001	1.09
App	2.49 (1.76)			
Browser	0.59 (0.57)			

[a] Box-Cox transformation was applied to engagement volume and engagement intensity.

Sex

The literature suggests that male and female users exhibit sizable differences in web-based engagement behaviors [17]. Therefore, sex may have some influence on eHealth website engagement. We performed a *t* test to compare the engagement volume and engagement intensity between female and male users, and the results are shown in Table 8.

Table 8. Comparison of user engagement between sexes[a].

User engagement and sexes	Value, mean (SD)	*t* test (*df*)	*P* value	Cohen *d*
Engagement volume		0.82 (371)	.41	0.10
Female	4.06 (2.46)			
Male	3.83 (2.15)			
Engagement intensity		−8.36 (371)	<.001	0.38
Female	2.14 (1.58)			
Male	2.84 (1.93)			

[a] Box-Cox transformation was applied to engagement volume and engagement intensity.

The results in Table 8 suggest that the difference in the engagement volume between female and male users is not significant (Cohen *d*=0.10; t_{371}=0.82; *P*=.41, medium effect size). However, male users had a greater engagement intensity (Cohen *d*=0.38; t_{371}=−8.36; *P*<.001, large effect size) than female users. One possible explanation is that male users are more proficient in information technology skills (eg, internet information retrieval skills) [28].

Income

The literature also suggests that high-income and low-income users exhibit some differences in engagement behavior [17]. Therefore, income is also identified as a potential variable that may influence eHealth website engagement. In this study, income is measured using a proxy variable, monthly telecom expenditures. This is because data on actual income are difficult to obtain, and users with high incomes usually correspond to users with high telecom expenditures. Therefore, we performed a *t* test to compare the engagement volume and engagement intensity between low- and high-income users, and the results are shown in Table 9.

Table 9. Comparison of user engagement between low- and high-income users[a].

User engagement and types of users	Value, mean (SD)	*t* test (*df*)	*P* value	Cohen *d*
Engagement volume		0.94 (371)	.35	0.11
Low-income users	4.03 (2.34)			
High-income users	3.79 (2.17)			
Engagement intensity		−1.07 (371)	.29	0.04
Low-income users	2.57 (1.78)			
High-income users	2.65 (1.85)			

[a] Box-Cox transformation was applied to engagement volume and engagement intensity.

The results in Table 9 suggest that the difference in the engagement volume between low- and high-income users is not significant (t_{371}=0.94; *P*=.35). Furthermore, the difference in engagement intensity between low- and high-income users was not significant (t_{371}=−1.07; *P*=.29). Therefore, we found no significant influence of income on the engagement patterns of eHealth websites.

Interaction Analysis

Interactions may exist among the four factors identified earlier. Therefore, an analysis of variance was conducted to test the potential interaction effects, and the results are shown in Table 10. The results in Table 10 indicate that the interaction between the platform and income is significant for both engagement volume ($F_{1,310}$=6.20; *P*=.01) and engagement intensity ($F_{1,1817}$=30.15; *P*<.001). This finding implies that although the main effect of income on engagement is not significant, it moderates the influence of the platform on engagement.

Table 10. The analysis of variance results[a].

Factor	*F* test (*df*)	*P* value
Engagement volume		
Platform×income [b]	*6.20 (1)*	*.01*
Channel×income	0.03 (1)	.86
Platform×sex	1.36 (1)	.24
Channel×sex	0.09 (1)	.76
Engagement intensity		
Platform×income	*30.15 (1)*	*<.001*
Channel×income	0.28 (1)	.60
Platform×sex	0.12 (1)	.73
Channel×sex	0.18 (1)	.67

[a] Box-Cox transformation was applied to engagement volume and engagement intensity.

[b] Italicization denotes significance (*P*<.05).

The details of the interaction between the platform and income are shown in Table 11. The results in Table 11 suggest that income negatively moderates the influence of the platform on engagement volume and engagement intensity. That is, the advantage of Android users over iOS users regarding engagement volume and engagement intensity is more salient among low-income users. More specifically, Android users have a significantly higher engagement volume and engagement intensity than iOS users when they are low income. However, Android users only have a significantly higher engagement intensity than iOS users when they are high income. For high-income users, the engagement volumes for Android users and iOS users were not significantly different. Low-income Android users are the users who are the most engaged in eHealth websites. Surprisingly, low-income iOS users are those who are least engaged in eHealth websites.

Table 11. The interaction between the platform and income[a].

Income and platform	Engagement volume	Engagement intensity
Low		
Android	4.55	2.86
iOS	3.22	1.36
High		
Android	3.97	2.41
iOS	3.98	1.63

[a] Box-Cox transformation was applied to engagement volume and engagement intensity.

Discussion

Principal Findings

Several major findings were obtained in this study. First, the market share analysis indicates that the medical category accounts for the largest market share of eHealth website visits (134,009/184,826, 72.51%), followed by the lifestyle category (46,870/184,826, 25.36%). The e-pharmacy category had the smallest market share, accounting for only 2.14% (3947/184,826) of the total visits.

Second, eHealth websites are characterized by very low visit penetration but relatively high user penetration. This means that although eHealth websites are associated with a low usage frequency, they are closely related to everyone and have great market potential.

Third, the distribution of engagement intensity follows a power law distribution. A large number of eHealth needs involve only a small number of visits, whereas a very small number of complex eHealth needs must be realized through a large number of visits.

Fourth, visits to eHealth websites were highly concentrated. Most users (238/373, 63.8%) visited only one eHealth website within 3 months. On average, each user visits 1.5 eHealth websites. Fewer than 40% of users visit multiple eHealth websites.

Fifth, there are day and hour trends in eHealth website engagement patterns. User engagement is generally high on weekdays but low on weekends. In addition, user engagement increases gradually from morning to noon. After noon, user engagement declines until it reaches its lowest level at midnight.

Sixth, customer loyalty also differed significantly among the categories. Lifestyle websites, followed by medical websites, had the highest customer loyalty. e-Pharmacy websites had the lowest customer loyalty.

Seventh, cross-site browsing among categories was not symmetrical. For example, 66% (31/47, 66%) of lifestyle website users visited medical websites, whereas only 25% (31/124, 25%) of medical website users visited lifestyle websites. The asymmetric nature indicates that popular eHealth websites, such as medical websites, can effectively provide referral traffic for lifestyle and e-pharmacy websites. However, the opposite is not true.

Eighth, Android users are more engaged than iOS users on eHealth websites. This is because users can find more health apps that cost less on the Android platform. Another possible explanation is that Android users are more introverted or more proficient in information technology.

Ninth, app users are much more engaged than browser users. The engagement volume of app users is 4.85 times that of browser users, and the engagement intensity of app users is 4.22 times that of browser users. Such a sizable engagement gap can be explained by the great advantage of apps over browsers.

Tenth, male users had greater engagement intensity than female users. The engagement gap between male and female users can be explained by the fact that male users are more proficient in information technology skills.

Finally, income negatively moderates the influence of the platform (Android vs iOS) on user engagement. The advantage of Android users over iOS users regarding engagement volume and engagement intensity is more salient among low-income users. Low-income Android users are the users most engaged on eHealth websites. Conversely, low-income iOS users are those who are least engaged on eHealth websites.

Limitations

This study also has some limitations. First, the sample size used in this study was not very large. Only 373 users from Shanghai, China, were included in the data set. More users should be incorporated in future analyses. Second, the income variable used in this study was measured using a proxy. It is measured by the monthly telecom expenditure. Although monthly expenditures should be associated with user income, their relationship is not deterministic. Better approaches, such as surveys, can be used to measure user income in future studies.

Conclusions

In this study, we provide an overview of user engagement behavior on eHealth websites based on cross-site clickstream data. More specifically, we conducted an analysis to determine the market shares of different categories of eHealth websites, penetration of eHealth behavior, engagement intensity, engagement variety, day and hour trends, customer loyalty, and cross-site engagement behavior. Furthermore, we investigated the factors that influence user engagement on eHealth websites. The results indicate that the platform (Android vs iOS), channel (browser vs app), and sex (female vs male) have significant influences on engagement behavior. In addition, income (high vs low) negatively moderates the influence of platforms on engagement behavior.

Future research may focus on how the configuration of eHealth website resources may influence user engagement. Each eHealth website may have some health care resources (eg, health information, e-consultation, provider rating, and web-based registration). According to the resource orchestration theory, the role of one resource is not independent. Instead, its effect depends on the presence of other resources. How the configuration of resources may influence user engagement is an important RQ for the managers of eHealth websites. A configurational approach (eg, fuzzy set qualitative comparative analysis) can be used in the future to investigate the best resource composition pattern for eHealth websites.

Acknowledgments

This research was supported by the Humanity and Social Science Youth Foundation of the Ministry of Education of China (grant 18YJC630068) and the National Natural Science Foundation of China (grants 71971082 and 71471064).

Authors' Contributions

JL and JY conceived the research and design of the study protocol. KY, XB, and XL performed the study and collected the data. All authors (JL, KY, XB, XL, and JY) contributed to drafting the manuscript.

Conflicts of Interest

None declared.

References

1. Wang L, Yan L, Zhou T, Guo X, Heim GR. Researching mental health disorders in the era of social media: systematic review. Inform Syst Res 2020 Jun;31(2):537-555. [doi: 10.1287/isre.2019.0901]
2. Ghose A, Goldfarb A, Han SP. How is the mobile internet different? Search costs and local activities. Inform Syst Res 2013 Sep;24(3):613-631. [doi: 10.1287/isre.1120.0453]
3. Wei Y, Zheng P, Deng H, Wang X, Li X, Fu H. Design features for improving mobile health intervention user engagement systematic review and thematic analysis. J Med Internet Res 2020;22(12):e21687. [doi: 10.2196/preprints.21687]
4. Aitken M, Clancy B, Nass D. The Growing Value of Digital Health. The IQVIA Institute. 2017. URL: https://www.iqvia.com/insights/the-iqvia-institute/reports/the-growing-value-of-digital-health [accessed 2021-02-07]
5. Böhm AK, Jensen ML, Sørensen MR, Stargardt T. Real-world evidence of user engagement with mobile health for diabetes management: longitudinal observational study. JMIR Mhealth Uhealth 2020 Nov 6;8(11):e22212 [FREE Full text] [doi: 10.2196/22212] [Medline: 32975198]
6. Glasgow RE, Christiansen SM, Kurz D, King DK, Woolley T, Faber AJ, et al. Engagement in a diabetes self-management website: usage patterns and generalizability of program use. J Med Internet Res 2011 Jan 25;13(1):e9 [FREE Full text] [doi: 10.2196/jmir.1391] [Medline: 21371992]
7. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
8. Rahman QA, Janmohamed T, Pirbaglou M, Ritvo P, Heffernan JM, Clarke H, et al. Patterns of user engagement with the mobile app, manage my pain: results of a data mining investigation. JMIR Mhealth Uhealth 2017 Jul 12;5(7):e96 [FREE Full text] [doi: 10.2196/mhealth.7871] [Medline: 28701291]
9. Zhang Y, Xia T, Huang L, Yin M, Sun M, Huang J, et al. Factors influencing user engagement of health information disseminated by Chinese provincial centers for disease control and prevention on wechat: observational study. JMIR Mhealth Uhealth 2019 Jun 27;7(6):e12245 [FREE Full text] [doi: 10.2196/12245] [Medline: 31250833]
10. 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]
11. Kashian N, Jacobson S. Factors of engagement and patient-reported outcomes in a stage IV breast cancer Facebook group. Health Commun 2020 Jan 23;35(1):75-82. [doi: 10.1080/10410236.2018.1536962] [Medline: 30351185]
12. Wang X, High A, Wang X, Zhao K. Predicting users' continued engagement in online health communities from the quantity and quality of received support. J Assoc Inf Sci Technol 2020 Dec 3;72(6):710-722. [doi: 10.1002/asi.24436]
13. Edney S, Ryan JC, Olds T, Monroe C, Fraysse F, Vandelanotte C, et al. User engagement and attrition in an app-based physical activity intervention: secondary analysis of a randomized controlled trial. J Med Internet Res 2019 Nov 27;21(11):e14645 [FREE Full text] [doi: 10.2196/14645] [Medline: 31774402]
14. Comello ML, Qian X, Deal AM, Ribisl KM, Linnan LA, Tate DF. Impact of game-inspired infographics on user engagement and information processing in an ehealth program. J Med Internet Res 2016 Sep 22;18(9):e237 [FREE Full text] [doi: 10.2196/jmir.5976]

15. Brusk JJ, Bensley RJ. A comparison of mobile and fixed device access on user engagement associated with women, infants, and children (WIC) online nutrition education. JMIR Res Protoc 2016 Nov 15;5(4):e216 [FREE Full text] [doi: 10.2196/resprot.6608] [Medline: 27847351]
16. Peng Z, Sun Y, Guo X. Antecedents of employees’ extended use of enterprise systems: an integrative view of person, environment, and technology. Int J Inform Manag 2018 Apr;39:104-120. [doi: 10.1016/j.ijinfomgt.2017.11.007]
17. Nadeem W, Andreini D, Salo J, Laukkanen T. Engaging consumers online through websites and social media: a gender study of Italian Generation Y clothing consumers. Int J Inform Manag 2015 Aug;35(4):432-442. [doi: 10.1016/j.ijinfomgt.2015.04.008]
18. Hill S, Jansen M. Android vs iOS: Which Smartphone Platform Is the Best? 2021. Digital Trends. URL: https://www.digitaltrends.com/mobile/android-vs-ios/ [accessed 2021-02-07]
19. Schneider JA, Holland CP. Ehealth search patterns: a comparison of private and public health care markets using online panel data. J Med Internet Res 2017 Apr 13;19(4):e117 [FREE Full text] [doi: 10.2196/jmir.6739] [Medline: 28408362]
20. Gillespie CS. Fitting heavy tailed distributions: the package. J Stat Soft 2015;64(2):1-16. [doi: 10.18637/jss.v064.i02]
21. Li Q, Guo X, Bai X. Weekdays or weekends: exploring the impacts of microblog posting patterns on gratification and addiction. Inform Manag 2017 Jul;54(5):613-624. [doi: 10.1016/j.im.2016.12.004]
22. Park Y, Fader PS. Modeling browsing behavior at multiple websites. Marketing Science 2004 Aug;23(3):280-303. [doi: 10.1287/mksc.1040.0050]
23. Nelson R. Global App Revenue Grew 23% in 2018 to More Than $71 Billion on iOS and Google Play. Sensortower. 2019. URL: https://sensortower.com/blog/app-revenue-and-downloads-2018 [accessed 2021-02-07]
24. Buildfire. Android vs iOS Users: How Do They Behave Differently? Build Fire. 2017. URL: https://buildfire.com/ios-android-users/ [accessed 2021-02-07]
25. What Your Choice of Smartphone Says About You. Society BP. 2016. URL: https://www.eurekalert.org/pub_releases/2016-08/bps-wyc083116.php [accessed 2021-02-07]
26. Fritz CO, Morris PE, Richler JJ. Effect size estimates: current use, calculations, and interpretation. J Exp Psychol Gen 2012 Feb;141(1):2-18. [doi: 10.1037/a0024338] [Medline: 21823805]
27. Sawilowsky SS. New effect size rules of thumb. J Mod App Stat Meth 2009 Nov 1;8(2):597-599. [doi: 10.22237/jmasm/1257035100]
28. Rowley J, Johnson F, Sbaffi L. Gender as an influencer of online health information-seeking and evaluation behavior. J Assn Inf Sci Tec 2015 Aug 25;68(1):36-47. [doi: 10.1002/asi.23597]

Abbreviations

RQ: research question

Edited by R Kukafka; submitted 01.04.21; peer-reviewed by D Gu, X Yang; comments to author 19.04.21; revised version received 28.04.21; accepted 24.05.21; published 13.08.21.

Please cite as:
Li J, Yu K, Bao X, Liu X, Yao J
Patterns of eHealth Website User Engagement Based on Cross-site Clickstream Data: Correlational Study
J Med Internet Res 2021;23(8):e29299
URL: https://www.jmir.org/2021/8/e29299
doi:10.2196/29299
PMID:34397392

©Jia Li, Kanghui Yu, Xinyu Bao, Xuan Liu, Junping Yao. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Review

Applications of Extended Reality in Ophthalmology: Systematic Review

Chee Wui Ong[1], MBBS; Marcus Chun Jin Tan[2], MBBS, MMed; Michael Lam[3], MBBS, MMed; Victor Teck Chang Koh[2,4], MBBS, MMed, MRCSEd

[1]Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
[2]Department of Ophthalmology, National University Hospital, Singapore, Singapore
[3]Department of Ophthalmology, Ng Teng Fong General Hospital, Singapore, Singapore
[4]Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

Corresponding Author:
Victor Teck Chang Koh, MBBS, MMed, MRCSEd
Department of Ophthalmology
Yong Loo Lin School of Medicine
National University of Singapore
1E Kent Ridge Road NUHS Tower Block, Level 7
Singapore, 119228
Singapore
Phone: 65 90230056
Email: victor_koh@nuhs.edu.sg

Abstract

Background: Virtual reality, augmented reality, and mixed reality make use of a variety of different software and hardware, but they share three main characteristics: immersion, presence, and interaction. The umbrella term for technologies with these characteristics is extended reality. The ability of extended reality to create environments that are otherwise impossible in the real world has practical implications in the medical discipline. In ophthalmology, virtual reality simulators have become increasingly popular as tools for surgical education. Recent developments have also explored diagnostic and therapeutic uses in ophthalmology.

Objective: This systematic review aims to identify and investigate the utility of extended reality in ophthalmic education, diagnostics, and therapeutics.

Methods: A literature search was conducted using PubMed, Embase, and Cochrane Register of Controlled Trials. Publications from January 1, 1956 to April 15, 2020 were included. Inclusion criteria were studies evaluating the use of extended reality in ophthalmic education, diagnostics, and therapeutics. Eligible studies were evaluated using the Oxford Centre for Evidence-Based Medicine levels of evidence. Relevant studies were also evaluated using a validity framework. Findings and relevant data from the studies were extracted, evaluated, and compared to determine the utility of extended reality in ophthalmology.

Results: We identified 12,490 unique records in our literature search; 87 met final eligibility criteria, comprising studies that evaluated the use of extended reality in education (n=54), diagnostics (n=5), and therapeutics (n=28). Of these, 79 studies (91%) achieved evidence levels in the range 2b to 4, indicating poor quality. Only 2 (9%) out of 22 relevant studies addressed all 5 sources of validity evidence. In education, we found that ophthalmic surgical simulators demonstrated efficacy and validity in improving surgical performance and reducing complication rates. Ophthalmoscopy simulators demonstrated efficacy and validity evidence in improving ophthalmoscopy skills in the clinical setting. In diagnostics, studies demonstrated proof-of-concept in presenting ocular imaging data on extended reality platforms and validity in assessing the function of patients with ophthalmic diseases. In therapeutics, heads-up surgical systems had similar complication rates, procedural success rates, and outcomes in comparison with conventional ophthalmic surgery.

Conclusions: Extended reality has promising areas of application in ophthalmology, but additional high-quality comparative studies are needed to assess their roles among incumbent methods of ophthalmic education, diagnostics, and therapeutics.

(J Med Internet Res 2021;23(8):e24152) doi:10.2196/24152

KEYWORDS

extended reality; virtual reality; augmented reality; mixed reality; ophthalmology; ophthalmic

Introduction

The rapid development of extended reality technologies has necessitated recent efforts to define and draw lines between new concepts and subgroups of extended reality applications [1]. Virtual reality has been defined as one in which our natural surroundings are completely replaced with a 3D computer-generated environment via wearable screens in the form of head-mounted displays [2]. Augmented reality is a superimposition of computer-generated content with limited interactivity onto our visible surroundings. Mixed reality is similar to augmented reality, except that the user is able to interact vividly with computer-generated content [1]. Mixed reality can be considered an amalgamation of the features of both virtual reality and augmented reality, as both highly interactive computer-generated objects and the real physical world are integrated to dynamically coexist within a single display [3,4]. Whereas virtual reality, augmented reality, and mixed reality make use of a variety of different software and hardware, these extended reality technologies share 3 main characteristics: immersion, presence, and interaction [2,5]. *Immersion* refers to a perception of physical existence within the extended reality environment, *presence* describes the perception of connection to the environment, whereas *interaction* is the ability to act and receive feedback within the environment [2].

In medicine, the nascent influence of extended reality is prevalent. Virtual reality platforms have been designed to teach foundational subjects, such as human anatomy [6,7], and train surgeons in complex surgical procedures [8-11]. Augmented and mixed reality offer methods of visualizing intraoperative procedures and diagnostic images with devices, such as Google Glass (Google Inc) or Microsoft HoloLens (Microsoft Inc), that have the potential to improve procedure safety and success [12-14]. The ability of virtual reality to distract patients from the physical environment also offers therapeutic approaches for rehabilitation and for treating pain or psychiatric disorders [15-17]. Likewise, ophthalmology has seen a growing influence of extended reality. Ophthalmic graduate medical education in the United States has seen an increase in the use of virtual eye surgery simulators, from 23% in 2010 to 73% in 2018 [18,19]. Extended reality technologies have also been explored as a method of therapy in ophthalmic diseases such as amblyopia and visual field defects [20,21]. Although the versatility of extended reality platforms can influence the practice of ophthalmology, health care providers should be well informed of the benefits and limitations of such technologies. This will allow evidence-based decision making when adopting nascent methods of ophthalmic education, diagnosis, and treatment. The focus of this review was to systematically evaluate current evidence of the efficacy, validity, and utility of the application of extended reality in ophthalmic education, diagnostics, and therapeutics.

Methods

Eligibility Criteria

We included studies evaluating the use of extended reality for ophthalmic applications in education, diagnostics, and therapeutics for eye care professionals and ophthalmic patients. All study designs were included with the exception of systematic reviews, case reports, and case series with ≤3 patients. Non-English publications and publications on the technical engineering of extended reality were excluded.

Search Methods

Three databases served as the source of our search—PubMed MEDLINE, Embase, and Cochrane Register of Controlled Trials. Search terms included “Virtual Reality,” “Augmented Reality,” “Mixed Reality,” “Simulation,” “Simulated,” “3D,” “Ophthalmology,” “Ophthalmic,” and “Eye.” The search was performed on April 15, 2020. Publications from January 1, 1956 to April 2020 were searched without language or publication-type restrictions. References in studies meeting the eligibility criteria were searched to identify additional eligible studies. EndNote X9 (2020; Clarivate Analytics) was used to manage all identified publications and remove duplicates (Multimedia Appendix 1). Search results were recorded according to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) guidelines [22].

Study Selection

Two authors (CWO and MCJT) read all titles returned by the search. All abstracts of relevant titles and full texts of the relevant abstracts were read by the same authors to evaluate eligibility. Any uncertainties was resolved by discussion among all authors.

Data Collection and Analysis

For each study that met eligibility criteria, the quality of study was evaluated using Oxford Centre for Evidence-Based Medicine (OCEBM) levels of evidence [23].

Information from each study was extracted, including aim, design, population, sample size, extended reality technology type, application, outcomes, and findings.

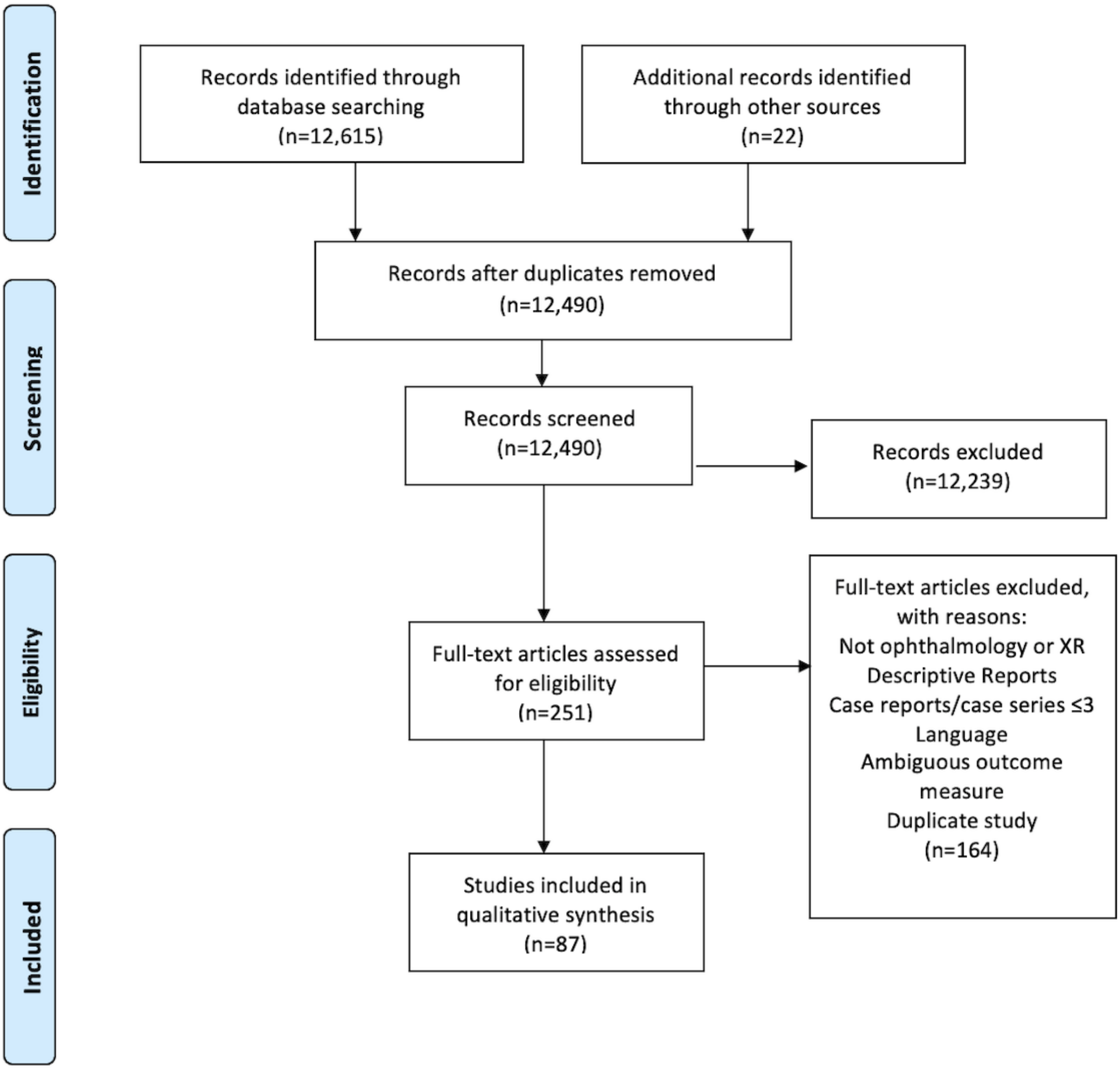
A number of eligible studies investigated the use of extended reality educational training simulators as training and assessment tools. Evidence of validity should be used to support the appropriateness of interpretation of results from assessments of performance using these simulators [24,25]. Validation is critical to be able to trust the results of a given education tool, and educators need evidence of validity to identify the appropriate assessment tool to meet specific educational needs with finite resources [26]. We chose a contemporary model of validity [24], comprising 5 sources of validity evidence—Content, Response process, Internal structure, Relationship to other variables, and Consequences [25,27] (Multimedia Appendix 2), to evaluate the extent to which the validity of these simulator-based assessments had been established by evidence. Due to a high degree of heterogeneity between studies, quantitative statistical analysis was not conducted.

Results

General

A total of 12,490 unique records were identified. After screening by title and abstract, 251 full-text publications were retrieved for assessment for final eligibility. Of these, 164 were excluded, and 87 studies met the final eligibility criteria (Figure 1). Of these, 54 were relevant to the use of extended reality in education, 5 were relevant to the use of extended reality in diagnostics, and 28 were relevant to the use of extended reality in therapeutics.

Figure 1. Flow Diagram showing inclusion process for identified records. XR: extended reality.


Identification
Records identified through database searching (n=12,615)
Additional records identified through other sources (n=22)
Records after duplicates removed (n=12,490)
Screening
Records screened (n=12,490)
Records excluded (n=12,239)
Eligibility
Full-text articles assessed for eligibility (n=251)
Full-text articles excluded, with reasons:
Not ophthalmology or XR
Descriptive Reports
Case reports/case series ≤3
Language
Ambiguous outcome measure
Duplicate study
(n=164)
Included
Studies included in qualitative synthesis (n=87)

Education

Overview

Applications of extended reality in education included surgical simulators (46/54), ophthalmoscopy simulators (6/54), and optometry training simulators (2/54), with medical students, optometry students, trainee, or trained ophthalmologists as participants.

Surgical Simulators

Of 46 studies evaluating surgical simulators, the EyeSi surgical simulator (VR Magic) was most commonly used (n=38). Others included MicroVisTouch (ImmersiveTouch) (n=1), PixEye Ophthalmic Simulator (SimEdge SA) (n=1), and 6 self-designed simulators. The most common surgical procedure simulated in these studies was cataract surgery (n=36), followed by vitreoretinal procedures (n=9), laser trabeculoplasty (n=1), and corneal laceration repair (n=1).

Of the 46 studies, 24 were studies that evaluated the efficacy of surgical simulators using evaluations of surgical performance on real patients, objective assessments by subspecialty experts, or simulator-based metrics as outcome measures. There were 14 [28-47] studies that used surgical performance on real patients as outcome measures, of which 4 were randomized trials [48-51] (Table 1); these randomized trials compared the use of virtual reality ophthalmic surgical simulators with conventional methods of surgical training. Objective assessment metrics of participants' surgical performance on real patients were evaluated by subspecialty experts. All studies [48,50,51], except one [49], showed that simulator-training resulted in surgical performance significantly superior to that of conventional training methods in terms of quality, time efficiency, efficacy, or complication rates. In particular, Deuchler et al [50] found that warm-up simulation training improved performance for surgeons who had not operated for a significant period of time. Daly et al [49] showed that residents who underwent EyeSi training were significantly slower at performing their first continuous curvilinear capsulorhexis than participants who underwent wet-lab training but achieved similar surgical performance scores.

Table 1. Randomized trials evaluating efficacy of surgical simulators in improving surgical performance on real patients.

Study	Design (OCEBM[a] level)	Intervention group training	Control group training	Simulated task	Population	Findings
Peugnet (1998) [48]	Randomized trial (2b)	Laser photocoagulation simulator (n=5)	Real patients (n=4)	Retinal photo-coagulation	Eye residents	Simulator training was significantly more time efficient (time efficiency index of 0.59 vs 0.28, *P*<.05), and resulted in a trend of greater photocoagulation efficiency (duration/impact index of 0.040 vs 0.028)
Daly (2013) [49]	Randomized trial (2b)	EyeSi (n=11)	Wet lab (n=10)	Continuous curvilinear capsulorhexis	Eye residents	Wet-lab trained residents significantly faster than EyeSi-trained residents (*P*=.038)
Deuchler (2016) [50]	Randomized trial (1b)	EyeSi warm-up (n=9)	No warm-up (n=12)	Pars plana vitrectomy	Surgeries by vitreoretinal surgeons	EyeSi warm-up improved surgical performance significantly (GRASIS[b] score of 1.0 vs 0.5, *P*=.0302)
Alwadani (2012) [51]	Randomized trial (2b)	PixEyes simulator (n=24)	Didactic, wet lab (n=23)	Argon laser trabeculoplasty	Medical students	VR-trained students had lower rates of inadvertent corneal/iris burns (4.5% vs 34.0%, *P*=.01), delivery misses (8% vs 55%, *P*=.001), overtreatment and undertreatment (7% vs 46%, *P*=.015)

[a] OCEBM: Oxford Centre for Evidence-Based Medicine.

There were 10 nonrandomized studies [28-37] that evaluated the EyeSi simulator for cataract surgery training (Table 2). Thomsen et al [28] and la Cour et al [29] demonstrated statistically significant improved Objective Structured Assessment of Cataract Surgical Skill Scores (OSACSS) in novice surgeons after training with the EyeSi cataract module. Roohipoor et al [30] found significant correlations between residents' EyeSi simulator-based scores and their eventual surgery count and Global Rating Assessment of Skills in Intraocular Surgery scores. The other 7 studies [31-37] investigated the relationship between EyeSi simulator use and complication rates in cataract surgery on real patients, of which 6 studies [31-36] showed that the use of the EyeSi simulator was associated with reduced complication rates. McCannel et al [37] found that EyeSi capsulorhexis training was not associated with lower vitreous loss rates overall but was associated with higher nonerrant continuous curvilinear capsulorhexis-associated vitreous loss.

Table 2. Nonrandomized trials evaluating efficacy of surgical simulators in improving surgical performance on real patients.

Study	Design (OCEBM[a] level)	Comparison or control group	Outcome measure	Participants and cataract surgeries, (n)	Findings
Thomsen (2017a) [28]	Cohort (2b)	Before EyeSi use	OSACSS[b]	Cataract surgeons (19)	Novices and less-experienced surgeons showed significant improvements in the operating room (32% and 38% improvement, *P*=.008 and *P*=.018 respectively) after EyeSi training
La Cour (2019) [29]	Cohort (2b)	Before EyeSi use	OSACSS	Cataract surgeons (19)	EyeSi training resulted in significantly improved surgical performance in less-experienced surgeons. Skill-transfer between modules was not demonstrable
Roohipoor (2017) [30]	Cohort (2b)	N/A[c]	GRASIS[d]	Ophthalmology residents (30)	Significant correlations between residents' EyeSi simulator-based scores and their eventual surgery count and GRASIS scores
Belyea (2011) [31]	Cohort (3b)	No EyeSi use	Phaco time, percentage power, complications	Surgeries by ophthalmology residents (592)	EyeSi training resulted in significantly lower procedure duration (*P*=.002), percentage power (*P*=.001), and nonsignificantly fewer intraoperative complications
Pokroy (2013) [32]	Cohort (2c)	No EyeSi use	Incidence of posterior capsule tears, operation duration	Surgeries by ophthalmology residents (1000)	EyeSi training resulted in nonsignificantly fewer posterior capsule tears and shorter learning curves
Ferris (2020) [33]	Cohort (2b)	No EyeSi access	Posterior capsule rupture rates	Surgeries by ophthalmology residents (17831)	Residents with EyeSi access had a significant reduction in posterior capsule rupture rates (4.2% vs 2.6%, Difference in Proportions 1.5%, 95% CI 0.5-2.6%, *P*=.003). Posterior capsule rupture rates significantly lower after access to EyeSi (3.5% to 2.6%, Difference in proportions 0.9%, 95% CI 0.4-1.5%, *P*=.001)
Lucas (2019) [34]	Cohort (2b)	No EyeSi use	Complication rates	Surgeries by ophthalmology residents (140)	EyeSi training resulted in significantly fewer complications (12.86 vs 27.14%, *P*=.031)
Staropoli (2018) [35]	Cohort (2b)	No EyeSi use	Complication rates	Surgeries by ophthalmology residents (955)	EyeSi training resulted in significantly fewer complications (2.4 vs 5.1%, *P*=.037)
McCannel (2013) [36]	Case series (4)	Reduced EyeSi use	Errant continuous curvilinear capsulorhexis rates	Surgeries by ophthalmology residents (1037)	EyeSi training resulted in significantly lower errant continuous curvilinear capsulorhexis rates (5.0 vs 15.7%, *P*<.001)
McCannel (2017) [37]	Case series (4)	Reduced EyeSi use	Vitreous loss rates, retained lens material	Surgeries by ophthalmology residents (1037)	EyeSi training was not associated with lower vitreous loss rates or less retained lens material but was associated with higher vitreous loss in nonerrant CCCs[e]

[a] OCEBM: Oxford Centre for Evidence-Based Medicine.

[b] OSACCS: Objective Structured Assessment of Cataract Surgical Skill Score.

[c] N/A: not applicable.

[d] GRASIS: Global Rating Assessment of Skills in Intraocular Surgery.

[e] CCC: continuous curvilinear capsulorhexis.

Seven studies [38-44] investigated the efficacy of the EyeSi surgical simulator (n=6) or a self-made augmented reality microsurgery simulator (n=1) by evaluating participants' surgical performance on the same simulators using simulator-based metrics (Table 3 and Table 4). Selvander et al [40] used the OSACSS and Objective Structured Assessment of Technical Surgical Skills (OSATS).

Three studies were randomized trials [38-40] (Table 3). Thomsen et al [39] investigated if there could be interprocedural transfer of skills and found that residents with simulated cataract surgery training did not perform significantly better than those without (simulator score with training: mean 381, SD 129 vs simulator score without training: mean 455, SD 82, *P*=.262) at the vitreoretinal surgery module. Selvander et al [40] had a similar aim and found that training on the capsulorhexis or cataract navigation training module on the EyeSi did not significantly improve performance at the other (OSACSS—training: 8, no training: 8, *P*=.64; OSATS score—training: 7, no training: 10, *P*=.52); however, repeated practice with each module significantly improved simulator-based scoring for the respective modules. Bergqvist et al [38] found that medical students who trained with simulated cataract surgery had higher overall simulator scores and fewer complications than those who did not train on the simulator. These trials suggest that simulator training improves performance only for the specific procedure being trained. The other 4 studies [41-44] were prospective cohort studies or case series (Table 4) that demonstrated that extended reality surgical training resulted in significant improvements on subsequent simulator-based performance scores.

Table 3. Randomized trials evaluating efficacy of surgical simulators in improving surgical performance as measured by the same simulator.

Study	Design (OCEBM[a] level)	Intervention group	Control group	Participants	Simulated task
Bergqvist (2014) [38]	Randomized trial (1b)	EyeSi training (n=10)	No EyeSi training (n=10)	Medical students	Cataract surgery
Thomsen (2017b) [39]	Randomized trial (2b)	EyeSi training (n=6)	No EyeSi use (n=6)	Eye residents	Cataract surgery, vitreoretinal surgery
Selvander (2012) [40]	Randomized trial (2b)	EyeSi cataract navigation training first (n=17)	EyeSi capsulorhexis training first (n=18)	Medical students	Capsulorhexis, cataract navigation

[a] OCEBM: Oxford Centre for Evidence-Based Medicine.

Table 4. Nonrandomized trials evaluating efficacy of surgical simulators in improving surgical performance as measured by the same simulator.

Study	Design (OCEBM[a] level)	Surgical simulator	Participants	Simulated task	Findings
Saleh (2013a) [41]	Case series (4)	EyeSi	Eye residents (n=17)	Cataract surgery	EyeSi training resulted in significantly improved scores (*P*<.001) among residents, especially for capsulorhexis and antitremor
Gonzalez-Gonzalez (2016) [42]	Case series (4)	EyeSi	Medical students, eye residents (n=14)	Capsulorhexis	EyeSi training resulted in significantly improved course scores for both dominant (33.4 vs. 46.5; *P*<.05) and nondominant hands (28.9 vs. 47.7; *P*<.001) and faster performance times (*P*<.001)
Bozkurt (2018) [43]	Cohort (2b)	EyeSi	Ophthalmologists (n=16)	Capsulorhexis module	EyeSi training resulted in significantly improved capsulorhexis scores (*P*=.001)
Ropelato (2020) [44]	Case series (4)	Microsoft HoloLens	Unspecified (n=47)	Internal limiting membrane peeling	There was significant improvement in micromanipulation performance scores after simulator-training

[a] OCEBM: Oxford Centre for Evidence-Based Medicine.

Three randomized trials investigated the effect of extended reality surgical simulator training on surgical performance in the wet lab (Table 5). Surgical performance was assessed using objective outcomes. Feldman et al [46] and Feudner et al [47] showed that training on the EyeSi simulator significantly improved wet-lab performance for corneal laceration repair and capsulorhexis, respectively, while Jonas et al [45] showed that training on a self-made virtual reality simulator improved wet-lab performance for pars plana vitrectomies.

Table 5. Studies evaluating efficacy of surgical simulators in improving surgical performance in the wet lab.

Study	Design (OCEBM[a] level)	Intervention group	Control group	Participants	Outcome measure tool
Jonas (2003) [45]	Randomized trial (2b)	Simulator training (n=7)	No simulator training (n=7)	Medical students, eye residents	Amount of vitreous removed, retinal lacerations, residual retinal detachment, duration
Feldman (2007) [46]	Randomized trial (2b)	EyeSi training (n=8)	No EyeSi training (n=8)	Medical students	Corneal Laceration Repair Assessment
Feudner (2009) [47]	Randomized trial (1b)	EyeSi training (n=31)	No EyeSi training (n=31)	Eye residents	Scoring based on capsulorhexis video

[a] OCEBM: Oxford Centre for Evidence-Based Medicine.

Of 46 studies, 20 studies evaluated the validity of surgical simulator-based assessments (Multimedia Appendix 3). Most validity studies achieved an OCEBM level of evidence of 2b, corresponding to exploratory cohort studies with good reference standards. The most common source of validity evidence was *Relationship with other variables*, addressed in 19 of 20 studies (95%). Studies achieved this by statistically evaluating the relationship between surgical performance on the simulator and participants' levels of expertise. *Content validity* was addressed in 18 studies (90%), *Response process* was addressed in 9 studies (45%), *Internal structure* was addressed in 5 studies (25%), and *Consequences* was addressed in 2 studies (10%). Only 2 of 20 (10%) studies addressed all 5 sources.

Of these 20 studies, 12 assessed surgical performance using simulator-based scoring only [39,43,52-61], 6 studies compared simulator-based scores with video-based scoring (OSACSS, OSATS, or motion-tracking software) [62-67], and 2 studies used video-based scoring only [40,68].

For the EyeSi surgical simulator, most studies found that the surgical performance of experienced participants was significantly better than that of less-experienced participants. Sikder et al [52] found that intervening surgical experience significantly improved capsulorhexis performance on the MicroVisTouch cataract surgery simulator. Lam et al [62] showed that in a self-made phacoemulsification simulator, more experienced participants attained significantly higher scores in all main procedures and completed tasks significantly faster.

Five studies [69-73] assessed the perception of ophthalmologists and medical students toward surgical simulators using user-reported outcome measures. These studies achieved OCEBM evidence levels of 4 (n=4) and 2b (n=1). Users found the EyeSi and a novel virtual reality continuous curvilinear capsulorhexis simulator to be useful in improving surgical skill, confidence, and understanding, while providing a safe and realistic alternative for training.

Ophthalmoscopy Simulators

Six studies [74-79] evaluated the use of extended reality as a tool for education in ophthalmoscopy. Simulators used were the EyeSi Augmented Reality Direct (n=1) and Binocular Indirect (n=3) ophthalmoscopy simulators, and 2 novel self-made direct ophthalmoscopy simulators comprising the HTC Vive Virtual Reality-Head-Mounted Display (n=1) and the RITECH II Virtual Reality-Head-Mounted Display (n=1).

Two randomized trials [74,75], with OCEBM evidence levels 2b, assessed the efficacy of the EyeSi Binocular Indirect Ophthalmoscopy simulator. Both studies showed that participants who trained with the EyeSi Binocular Indirect Ophthalmoscopy simulator performed significantly better than participants who underwent conventional training.

Three studies [75-77], with OCEBM evidence levels of 2b, assessed the validity of the EyeSi Binocular Indirect Ophthalmoscopy simulator (n=2) and the EyeSi Binocular Direct Ophthalmoscopy simulator (n=1) for training and assessment. All studies demonstrated *Relationships with other variables* as a source of validity evidence and found that participants with more experience had significantly higher ophthalmoscopy evaluation scores. *Content validity* was addressed in all studies. Only 1 study [77] addressed *Internal structure* by evaluating internal consistency between simulator modules and evaluated *Consequences* by calculating a pass or fail score.

Two user perception studies [78,79] found that medical students felt that self-assembled virtual reality direct ophthalmoscopy simulators were usable and useful in improving ophthalmoscopy skills.

Optometry Training Simulators

Two studies [80,81] evaluated the preliminary user experience of an augmented reality optometry simulator comprising a head-mounted display, a slit-lamp instrument, and a simulated eye, which allowed the simulation of optometry training tasks. User studies involving undergraduate optometry students showed that the simulator was feasible in simulating foreign body removal as a training task with a high level of user satisfaction.

Diagnostics

Overview

Five studies evaluated the use of extended reality for the production of immersive and interactive content for diagnostic applications. Two studies evaluated the use of extended reality to display ocular imaging data [82,83], and 3 studies [84-86] evaluated the validity of extended reality as a simulation tool for the functional assessment of patients with ophthalmic diseases.

Ocular Imaging

Two case series, which achieved OCEBM evidence levels of 4, evaluated the presentation of ocular imaging modalities in virtual reality and augmented reality environments.

Maloca et al [82] tested the feasibility of displaying optical coherence tomography images in a virtual reality environment with a virtual reality-head-mounted display. A user perception survey involving 57 participants found it to be well tolerated with minimal side-effects. Berger et al [83] demonstrated feasibility for a method of direct overlay of photographic and angiographic fundus images onto a real-time slit lamp fundus view in 5 participants.

Simulators for Functional Assessment

Three studies evaluated the use of extended reality simulators for the functional assessment of patients with ophthalmic diseases. The studies achieved OCEBM levels of evidence of 4.

Goh et al [84] trialed the use of the Virtual Reality Glaucoma Visual Function Test with a smartphone paired with the Google Cardboard head-mounted display to assess the visual function of glaucoma patients and found that stationary test person scores demonstrated criterion and convergent validity, corresponding to *Relationship with other variables*.

Ungewiss et al [85] compared the assessment of driving performance in a driving simulator with that in a real vehicle in patients with glaucoma (n=10), hemianopia (n=10), and normal controls (n=20) and found that patients with hemianopic and glaucoma performed worse than healthy controls on the driving simulator, demonstrating *Relationship with other variables* as a source of validity evidence.

Jones et al [86] evaluated the use of a head-mounted display to simulate visual impairment in glaucoma using virtual reality and augmented reality and found it able to replicate and objectively quantify functional impairments associated with visual impairments. When the simulated visual field loss was inferior, impairments were noted to be significantly greater than those noted when the simulated visual field loss was superior, which was consistent with previous experiences of real patients with glaucoma [87-89].

Therapeutics

Overview

A total of 28 studies evaluated the use of extended reality in therapeutics. These studies evaluated heads-up surgery (n=21), binocular treatment of amblyopia (n=2), functional improvement for the visually impaired (n=4), and an aid for achromatopsia (n=1).

Heads-up Surgery

Heads-up surgery involves the use of a 3D camera to capture images from a stereomicroscope for presentation on a 3D display. Of 21 studies [90-110], the most common heads-up surgical system evaluated was NGENUITY 3D (Alcon Laboratories) (n=12), followed by TrueVision 3D HD System (TrueVision Systems Inc) (n=2), TRENION 3D HD (Carl Zeiss Meditec) (n=1), MKC-700HD and CFA-3DL1 (Ikegami) (n=1), Digital Microsurgical Workstation (3D Vision Systems) (n=1), TIPCAM 1S 3D ORL endoscope (Karl Storz) (n=1). Surgical procedures included vitreoretinal procedures (n=17), cataract surgery (n=5), scleral buckle (n=1), and endoscopic lacrimal surgery (n=1).

Six studies had OCEBM evidence levels of 2b, corresponding to randomized trials (n=4) and cohort studies (n=2). There were 15 studies with OCEBM evidence levels of 4, corresponding to case series, case-control studies, or poor-quality cohort studies.

The 4 randomized trials [90-93] demonstrated noninferiority of heads-up surgery in comparison with conventional microscope surgery in postoperative outcomes and complications. Qian et al [90] performed phacoemulsification and intraocular lens implantation and reported no significant difference in mean surgery time, postoperative mean endothelial cell density between conventional surgery (n=10) and heads-up surgery (n=10). Talcott et al [91] performed pars plana vitrectomies and showed that compared with conventional surgery (n=16), heads-up surgery (n=23) significantly increased macular peel time (14.76 minutes vs 11.87 minutes, P=.004) but not overall operative time. There was no significant difference in visual acuity (logarithm of the minimum angle of resolution) or change from baseline, and no clinically significant intraoperative complications. Romano et al [92] randomized 50 eyes to the use of an unspecified heads-up surgical system (n=25) and conventional surgery (n=25) for 25-gauge pars plana vitrectomy; there was no significant difference in mean operation duration. Surgeons and observers were significantly more satisfied (P<.001) using the heads-up system. Kumar et al [93] randomized 50 patients to macular hole surgery with an unspecified heads-up system (n=25) and conventional macular hole surgery (n=25); there were no significant differences in postoperative visual acuity, macule hole indices, surgical time, total internal limiting membrane peel time, number of flap initiations, and macular hole closure rates. Microscope illumination intensity (heads-up: 100%; conventional: 45%) and endoillumination was significantly lower in heads-up surgery (heads-up: 40%; conventional: 13%).

The other 17 studies [94-110] were case series or cohort studies (Multimedia Appendix 4). They reported experiences with different heads-up surgical systems with various surgical procedures and assessed a few common outcomes with the following findings.

Of studies comparing heads-up surgical systems with conventional surgery, most reported procedural success or success that was not significantly different from that of conventional surgery. There were no significant differences in mean procedural durations, postoperative visual acuity, or improvements in visual acuity. The minimum required endoillumination was lower in heads-up surgery than that in conventional surgery. Zhang et al [94] found a significantly lower mean minimum required endoillumination for heads-up surgery than that for conventional surgery (10% vs 35%, 598.7 vs 1913 lx, P<.001). Matsumoto et al [95] operated safely and successfully on 74 eyes with an endoillumination intensity of 3% using a heads-up system. No study reported major perioperative complications or significant differences in complication rates between heads-up surgery and conventional surgery. Users preferred heads-up surgery to conventional surgery.

Binocular Treatment of Amblyopia

Two studies [21,111] evaluated the efficacy of the use of extended reality for interactive and immersive binocular treatment for amblyopia. Lee et al [111] randomized 22 children with amblyopia (mean age 8.7 years, SD 1.3) to treatment with virtual reality videogaming on an unspecified virtual reality-head-mounted display (n=7), virtual reality videogaming and Bangerter foil (n=5), or Bangerter foil only (n=10), achieving an OCEBM evidence level of 2b. Two of 7 (29%) of patients in the virtual reality videogaming group, and 2 of 5 (40%) patients in the virtual reality videogaming with Bangerter foil group gained more than 0.2 in logarithm of the minimum angle of resolution of vision. Ziak et al [21] trialed the use of the Oculus Rift virtual reality head-mounted display for dichoptic virtual reality video gaming in treating 17 adults with amblyopia, achieving an OCEBM evidence level of 4. There was a significant improvement in mean amblyopic eye visual acuity (logarithm of the minimum angle of resolution: from mean 0.58, SD 0.35 to mean 0.43, SD 0.38; *P*<.01). Mean stereoacuity also improved significantly from 263.3 s of arc to 176.7 s of arc (*P*<.01).

Functional Improvement for Visual Disorders

Five case series studies [20,112-115] evaluated the use of extended reality technologies to improve the function of patients with visual impairment and disorders. All reached an OCEBM evidence level of 4.

Two studies evaluated the use of digital spectacles (DSpecs) to improve mobility in a total of 43 patients with peripheral visual field deficits [20,112]. DSpecs work by capturing, relocating, and resizing video signals to fit within a patient's visual field in real time using augmented reality technology. DSpecs enabled patients with visual field defects to have improved object identification, hand-eye coordination, and walking mobility.

Sanchez et al [113] evaluated Augmented Reality Tags for Assisting the Blind, an augmented reality system which helps the user determine the position of indoor objects by generating an audio-based representation of space. Blind participants perceived Augmented Reality Tags for Assisting the Blind to be a useful tool for assisting orientation and mobility tasks.

Tobler-Ammann et al [114] evaluated the use of virtual reality exercise games to encourage exploration of neglected space in patients with visuospatial neglect after stroke. Cognitive and spatial exploration skills trended toward improvement after the use of virtual reality exercise games and continued improving at follow-up in 5 of 7 participants. Adherence rates were high, and there were no adverse events.

Melillo et al [115] evaluated the efficacy of an augmented reality wearable improved vision system for patients with color vision deficiency. The system captures and remaps colors from the environment and displays it to the user via a head-mounted display. The system significantly improved Ishihara Vision Test scores in participants with color vision deficiency (mean score 5.8 vs 14.8, *P*=.03).

Discussion

Although a wide range of clinically evaluated ophthalmic applications of extended reality were identified, we predominantly focused on the following domains: education, diagnostics, and therapeutics. In education, simulators demonstrated efficacy and validity in improving surgical and ophthalmoscopy skills. In diagnostics, extended reality devices demonstrated proof-of-concept in displaying ocular imaging data and validity in assessing the function of patients with glaucoma. In therapeutics, heads-up surgical systems were found to be efficacious and safe alternatives to conventional microscope surgery. The overall evidence, however, for the utility of these applications is limited. Only 8 of 87 (9%) studies had OCEBM levels of evidence of 1b, which represented randomized trials with a narrow confidence interval, while 79 of 87 (91%) studies had OCEBM levels of evidence ranging from 2b to 4 (cohort studies, case-control studies, and case series). For extended reality applications only evaluated by 1 or 2 studies, this limited evidence makes it difficult to extrapolate their utility in a wider context.

The acquisition of motor skills, as described by the Fitts–Posner 3 - Stage Theory [116] and the Dreyfus and Dreyfus model [117], progresses from an initial cognitive phase, in which learners attempt to understand, to an associative phase, in which they modify movement strategies based on feedback, to a final autonomous phase, in which motor performance is fluid. Advances in surgical simulation technology have provided options to augment or replace traditional methods without compromising training efficiency and patient safety. Although most randomized trials [38,39,45-47] showed that simulator-trained groups performed better than those who had no training at all, only 3 studies [48,49,51] drew comparisons between simulator-trained participants and participants trained with conventional methods of surgical training such as wet-lab materials and real patients. The performance of simulator-trained inexperienced surgeons was superior to that of conventionally trained inexperienced surgeons only in the laser-based procedures [48,51]. Where intraocular surgery was concerned, the performance of simulator-trained inexperienced surgeons was comparable, but not superior, to that of wet-lab trained inexperienced surgeons. Moreover, they required a longer duration to complete the procedures [49]. This could be attributed to the fact that the instruments used in wet-lab training were identical to those used in the operating room, whereas surgical simulators may not have been able to emulate all tactile and ergonomic aspects of ophthalmic surgery. This suggests a continued role for wet-lab training, at least until simulators can closely replicate the full surgical experience.

It has been shown that practicing complex motor task skills is most effective in multiple short training sessions spaced over time with variable tasks [118]. Simulation-based training allows for this, and most efficacy studies [37,39,40] demonstrate that extended reality surgical simulators were able to improve surgical performance in the specific procedure for which the participants were trained; however, the same studies found little evidence of a crossover effect—that these improvements were applicable to other surgical procedures—suggesting that simulator training is highly specific. It is possible that intense focus on solitary surgical steps can result in lack of skill development in others [37], and the transfer of surgical skills thus cannot be anticipated when planning surgical training curriculum. Nonetheless, simulation training appears to reduce complications for both surgically naive and less experienced surgeons and to improve performance for experienced ones who have had a hiatus in undertaking procedures.

Studies evaluating ophthalmoscopy simulators found that simulator training can improve both direct and indirect ophthalmoscopy skills. In comparison with surgical simulators, ophthalmoscopy simulators are not as widely adopted in ophthalmic training curricula. One possible reason lies in the nature of the simulated task. Surgery performed by inexperienced novices risks harming the patient, whereas ophthalmoscopy constitutes a minor inconvenience to a patient in terms of discomfort and time. The availability of surgical simulators with efficacy and validity may be more of a necessity than ophthalmoscopy simulators.

Validity is the cornerstone upon which educational assessments depend to be appropriately justified in their application and without which the purpose of assessments in education will have little intrinsic meaning [25]. Most validity studies on educational simulators only addressed 1 or 2 sources of validity evidence. The presence of more studies addressing all sources of validity evidence would facilitate more robust interpretation of assessment scores in ophthalmic training.

In diagnostics, visualizing ocular imaging data, such as optical coherence tomography, fundus photography, and angiography in virtual reality and augmented reality can reveal important intraocular spatial relationships [119,120] and allow for interactive exploration of imaging data to aid education, understanding of diseases, clinical assessment, and therapy. These studies [119,120] demonstrated proof-of-concept, but more studies are needed to evaluate their efficacy and accuracy for clinical use. Extended reality applications also demonstrated validity evidence and feasibility in objectively assessing functional limitation and driving performance of glaucoma and hemianopic patients. The scope of their application, however, is currently limited by the small number of studies and low number of sources of validity evidence. While extended reality was able to simulate the visual environment, it was unable to account for nonvisual cues, such as sound and touch, that patients with ophthalmic disease might rely upon in daily function.

In therapeutics, heads-up surgery allowed for better visualization, better ergonomics, and reduced endoillumination intensities than those in traditional microscope surgery without compromising outcomes. Widespread adoption of heads-up surgery, however, is limited by a few factors. First, the comfort of assistant surgeons and anesthetists has been shown to be reduced due to the positioning of the heads-up display [94,96]. Second, the learning curve of heads-up surgery has yet to be studied comprehensively. Talcott et al [91] reported that surgeons had higher ease of use with the traditional microscope than with the NGENUITY 3D visualization system, showing that the preference for heads-up surgery was not unanimous. Additional experience with heads-up displays can guide ophthalmic surgeons in transitioning from traditional microscopes to these novel systems.

It has been shown that nonstereoscopic, nonimmersive binocular treatment is a promising approach in treating children with amblyopia, with positive outcomes in amblyopic eye visual acuity and stereoacuity [121,122]. Likewise, in our review, we found that stereoscopic immersive dichoptic stimulation conferred the same benefits onto amblyopic patients with amblyopia. The 2 studies included in our review reported high adherence rates [21,111], while there have been studies on nonimmersive dichoptic stimulation reporting lower adherence rates [123,124]. Although there has been no study comparing immersive dichoptic stimulation with nonimmersive dichoptic stimulation, we postulate that immersive dichoptic stimulation can engender better patient adherence and adherence to binocular treatment. Before immersive binocular treatment can be recommended over standard binocular treatment or even over conventional occlusion therapy, additional comparative studies are needed to determine if they would be appropriate replacements or adjuncts to conventional therapy. A cost-benefit analysis would also be important, given that conventional therapy is affordable yet still efficacious.

Extended reality applications are not without adverse effects. Studies have shown that viewing of 3D displays can induce objective changes to accommodative function, convergence, refractive errors, and tear films [125-130] and subjective symptoms such as asthenopia, motion sickness, fatigue, and head or neck discomfort [131]. Techniques such as discrete viewpoint control have been shown to potentially ameliorate these adverse effects, but they are not yet widely adopted [132]. Most studies in our review did not evaluate the incidence of adverse effects induced by the extended reality set-ups. While there is growing anticipation for the adoption of extended reality, more research is needed to ascertain if these adverse effects will significantly affect the efficacy of ophthalmic applications and shape user safety guidelines.

The cost of extended reality technologies will also be a major concern for potential users. One study [133] in 2013 estimated that the EyeSi surgical simulator would save the average US ophthalmic residency program $4980 yearly in nonsupply costs based on time saved in the operating room, requiring 34 years to recoup the simulator's cost price. Another study [134] in 2013 found that nonsupply cost savings from EyeSi use were higher in larger residency programs, but still insufficient to recoup costs at 10 years. These cost-analyses, however, do not make comparisons with conventional methods of ophthalmic surgical training. The ability of extended reality surgical simulators to simulate surgical scenarios that are otherwise impossible to replicate in a wet lab, such as posterior polar cataracts, specific clock hours of zonulysis, or a shallow anterior chamber, may represent intangible cost-savings in ophthalmic pedagogy with respect to additional time spent supervising surgeons and operating room staff, resources, and schedule. The availability of such comparisons might help to better define the role of an extended reality simulator in surgical training from the perspective of cost.

Extended reality promises utility in many areas of application by overcoming the limits of the unalterable physical environment. In ophthalmic surgical education, extended reality surgical simulators demonstrate efficacy and validity in improving surgical performance. Before surgical simulators can be considered to be a competitive alternative to traditional ophthalmic surgical training, 2 main barriers need to be addressed—cost and the need for additional high-quality comparative studies. Until these issues are addressed, surgical simulators can only play a supporting role in surgical training programs, despite their versatility and ability to provide quantitative feedback. In therapy, extended reality heads-up surgical systems have already seen popular use in ophthalmic surgery, with the literature showing that this type of system provides an efficacious and safe platform for surgical visualization. Other diagnostic and therapeutic applications mainly demonstrate proof-of-concept, with a lack of robust comparative evidence. Additional comparative studies with designs that allow a high level of evidence should be encouraged to explore the efficacy of extended reality in these varied ophthalmic applications. As extended reality is a nascent technology, we predict that it will only continue to demonstrate value and offer novel alternatives in ophthalmic education, diagnostics, and therapy.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 14 KB - jmir_v23i8e24152_app1.docx]

Multimedia Appendix 2
Messick’s five sources of validity evidence.
[DOCX File , 15 KB - jmir_v23i8e24152_app2.docx]

Multimedia Appendix 3
Studies evaluating validity of assessment based on surgical simulators.
[DOCX File , 18 KB - jmir_v23i8e24152_app3.docx]

Multimedia Appendix 4
Case-series and cohort studies evaluating the use of heads-up surgical systems.
[DOCX File , 17 KB - jmir_v23i8e24152_app4.docx]

References

1. Brigham TJ. Reality check: basics of augmented, virtual, and mixed reality. Med Ref Serv Q 2017;36(2):171-178. [doi: 10.1080/02763869.2017.1293987] [Medline: 28453428]
2. Yeung AWK, Tosevska A, Klager E, Eibensteiner F, Laxar D, Stoyanov J, et al. Virtual and augmented reality applications in medicine: analysis of the scientific literature. J Med Internet Res 2021 Feb 10;23(2):e25499 [FREE Full text] [doi: 10.2196/25499] [Medline: 33565986]
3. Milgram P, Kishino F. A taxonomy of mixed reality visual displays. IEICE Transactions on Information Systems 1994;E77D(12):1321-1329 [FREE Full text]
4. Martin G, Koizia L, Kooner A, Cafferkey J, Ross C, Purkayastha S, PanSurg Collaborative. Use of the HoloLens2 mixed reality headset for protecting health care workers during the covid-19 pandemic: prospective, observational evaluation. J Med Internet Res 2020 Aug 14;22(8):e21486-e21486 [FREE Full text] [doi: 10.2196/21486] [Medline: 32730222]
5. Cipresso P, Giglioli IAC, Raya MA, Riva G. The past, present, and future of virtual and augmented reality research: a network and cluster analysis of the literature. Front Psychol 2018 Nov 6;9:2086-2020 [FREE Full text] [doi: 10.3389/fpsyg.2018.02086] [Medline: 30459681]
6. Andrews C, Southworth MK, Silva JNA, Silva JR. Extended reality in medical practice. Curr Treat Options Cardiovasc Med 2019 Mar 30;21(4):18 [FREE Full text] [doi: 10.1007/s11936-019-0722-7] [Medline: 30929093]
7. Zhao J, Xu X, Jiang H, Ding Y. The effectiveness of virtual reality-based technology on anatomy teaching: a meta-analysis of randomized controlled studies. BMC Med Educ 2020 Apr 25;20(1):127 [FREE Full text] [doi: 10.1186/s12909-020-1994-z] [Medline: 32334594]
8. Mazur T, Mansour TR, Mugge L, Medhkour A. Virtual reality-based simulators for cranial tumor surgery: a systematic review. World Neurosurg 2018 Feb;110:414-422. [doi: 10.1016/j.wneu.2017.11.132] [Medline: 29198889]
9. Bric JD, Lumbard DC, Frelich MJ, Gould JC. Current state of virtual reality simulation in robotic surgery training: a review. Surg Endosc 2016 Jun;30(6):2169-2178. [doi: 10.1007/s00464-015-4517-y] [Medline: 26304107]

10. Alaker M, Wynn GR, Arulampalam T. Int J Surg 2016 May;29:85-94 [FREE Full text] [doi: 10.1016/j.ijsu.2016.03.034] [Medline: 26992652]
11. 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]
12. Jiang T, Yu D, Wang Y, Zan T, Wang S, Li Q. HoloLens-based vascular localization system: precision evaluation study with a three-dimensional printed model. J Med Internet Res 2020 Apr 17;22(4):e16852 [FREE Full text] [doi: 10.2196/16852] [Medline: 32301738]
13. Wei NJ, Dougherty B, Myers A, Badawy SM. Using Google Glass in surgical settings: systematic review. JMIR Mhealth Uhealth 2018 Mar 06;6(3):e54 [FREE Full text] [doi: 10.2196/mhealth.9409] [Medline: 29510969]
14. Eckert M, Volmerg JS, Friedrich CM. Augmented reality in medicine: systematic and bibliographic review. JMIR Mhealth Uhealth 2019 Apr 26;7(4):e10967 [FREE Full text] [doi: 10.2196/10967] [Medline: 31025950]
15. 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]
16. Parsons TD, Riva G, Parsons S, Mantovani F, Newbutt N, Lin L, et al. Virtual reality in pediatric psychology. Pediatrics 2017 Nov;140(Suppl 2):S86-S91 [FREE Full text] [doi: 10.1542/peds.2016-1758I] [Medline: 29093039]
17. de Rooij IJM, van de Port IGL, Meijer JG. Effect of virtual reality training on balance and gait ability in patients with stroke: systematic review and meta-analysis. Phys Ther 2016 Dec;96(12):1905-1918. [doi: 10.2522/ptj.20160054] [Medline: 27174255]
18. Ahmed Y, Scott IU, Greenberg PB. A survey of the role of virtual surgery simulators in ophthalmic graduate medical education. Graefes Arch Clin Exp Ophthalmol 2011 Aug 8;249(8):1263-1265. [doi: 10.1007/s00417-010-1537-0] [Medline: 20931214]
19. Paul SK, Clark MA, Scott IU, Greenberg PB. Virtual eye surgery training in ophthalmic graduate medical education. Can J Ophthalmol 2018 Dec;53(6):e218-e220. [doi: 10.1016/j.jcjo.2018.03.018] [Medline: 30502995]
20. Sayed AM, Kashem R, Abdel-Mottaleb M, Roongpoovapatr V, Eleiwa TK, Abdel-Mottaleb M, et al. Toward improving the mobility of patients with peripheral visual field defects with novel digital spectacles. Am J Ophthalmol 2020 Feb;210:136-145. [doi: 10.1016/j.ajo.2019.10.005] [Medline: 31606442]
21. Žiak P, Holm A, Halička J, Mojžiš P, Piñero DP. Amblyopia treatment of adults with dichoptic training using the virtual reality oculus rift head mounted display: preliminary results. BMC Ophthalmol 2017 Jun 28;17(1):105 [FREE Full text] [doi: 10.1186/s12886-017-0501-8] [Medline: 28659140]
22. 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 [FREE Full text] [doi: 10.7326/0003-4819-151-4-200908180-00135] [Medline: 19622511]
23. Jeremy HI, Paul G, Trish G, Carl H, Alessandro L, Ivan M, et al. The Oxford levels of evidence 2. Oxford Centre for Evidence-Based Medicine. URL: https://www.cebm.net/index.aspx?o=56532016 [accessed 2020-05-01]
24. Messick S. Foundations of validity: meaning and consequences in psychological assessment. ETS Research Report Series 2014 Aug 08;1993(2):i-18. [doi: 10.1002/j.2333-8504.1993.tb01562.x]
25. Downing SM. Validity: on meaningful interpretation of assessment data. Med Educ 2003 Sep;37(9):830-837. [doi: 10.1046/j.1365-2923.2003.01594.x] [Medline: 14506816]
26. Cook DA, Hatala R. Validation of educational assessments: a primer for simulation and beyond. Adv Simul (Lond) 2016;1:31 [FREE Full text] [doi: 10.1186/s41077-016-0033-y] [Medline: 29450000]
27. Ghaderi I, Manji F, Park YS, Juul D, Ott M, Harris I, et al. Technical skills assessment toolbox: a review using the unitary framework of validity. Ann Surg 2015 Feb;261(2):251-262. [doi: 10.1097/SLA.0000000000000520] [Medline: 24424150]
28. Thomsen ASS, Bach-Holm D, Kjærbo H, Højgaard-Olsen K, Subhi Y, Saleh GM, et al. Operating room performance improves after proficiency-based virtual reality cataract surgery training. Ophthalmology 2017 Apr;124(4):524-531. [doi: 10.1016/j.ophtha.2016.11.015] [Medline: 28017423]
29. la Cour M, Thomsen ASS, Alberti M, Konge L. Simulators in the training of surgeons: is it worth the investment in money and time? 2018 Jules Gonin lecture of the Retina Research Foundation. Graefes Arch Clin Exp Ophthalmol 2019 May;257(5):877-881. [doi: 10.1007/s00417-019-04244-y] [Medline: 30648208]
30. Roohipoor R, Yaseri M, Teymourpour A, Kloek C, Miller JB, Loewenstein JI. Early performance on an eye surgery simulator predicts subsequent resident surgical performance. J Surg Educ 2017;74(6):1105-1115. [doi: 10.1016/j.jsurg.2017.04.002] [Medline: 28434885]
31. Belyea DA, Brown SE, Rajjoub LZ. Influence of surgery simulator training on ophthalmology resident phacoemulsification performance. J Cataract Refract Surg 2011 Oct;37(10):1756-1761. [doi: 10.1016/j.jcrs.2011.04.032] [Medline: 21840683]
32. Pokroy R, Du E, Alzaga A, Khodadadeh S, Steen D, Bachynski B, et al. Impact of simulator training on resident cataract surgery. Graefes Arch Clin Exp Ophthalmol 2013 Mar;251(3):777-781. [doi: 10.1007/s00417-012-2160-z] [Medline: 23007233]
33. Ferris JD, Donachie PH, Johnston RL, Barnes B, Olaitan M, Sparrow JM. Royal College of Ophthalmologists' National Ophthalmology Database study of cataract surgery: report 6. the impact of EyeSi virtual reality training on complications

rates of cataract surgery performed by first and second year trainees. Br J Ophthalmol 2020 Mar;104(3):324-329. [doi: 10.1136/bjophthalmol-2018-313817] [Medline: 31142463]

34. Lucas L, Schellini SA, Lottelli AC. Complications in the first 10 phacoemulsification cataract surgeries with and without prior simulator training. Arq Bras Oftalmol 2019;82(4):289-294 [FREE Full text] [doi: 10.5935/0004-2749.20190057] [Medline: 30970123]
35. Staropoli PC, Gregori NZ, Junk AK, Galor A, Goldhardt R, Goldhagen BE, et al. Surgical simulation training reduces intraoperative cataract surgery complications among residents. Simul Healthc 2018 Feb;13(1):11-15 [FREE Full text] [doi: 10.1097/SIH.0000000000000255] [Medline: 29023268]
36. McCannel CA, Reed DC, Goldman DR. Ophthalmic surgery simulator training improves resident performance of capsulorhexis in the operating room. Ophthalmology 2013 Dec;120(12):2456-2461. [doi: 10.1016/j.ophtha.2013.05.003] [Medline: 23796766]
37. McCannel CA. Continuous curvilinear capsulorhexis training and non-rhexis related vitreous loss: the specificity of virtual reality simulator surgical training (an American Ophthalmological Society thesis). Trans Am Ophthalmol Soc 2017 Aug;115:T2 [FREE Full text] [Medline: 29021716]
38. Bergqvist J, Person A, Vestergaard A, Grauslund J. Establishment of a validated training programme on the Eyesi cataract simulator. a prospective randomized study. Acta Ophthalmol 2014 Nov;92(7):629-634 [FREE Full text] [doi: 10.1111/aos.12383] [Medline: 24612448]
39. Thomsen ASS, Kiilgaard JF, la Cour M, Brydges R, Konge L. Is there inter-procedural transfer of skills in intraocular surgery? a randomized controlled trial. Acta Ophthalmol 2017 Dec;95(8):845-851 [FREE Full text] [doi: 10.1111/aos.13434] [Medline: 28371367]
40. Selvander M, Åsman P. Virtual reality cataract surgery training: learning curves and concurrent validity. Acta Ophthalmol 2012 Aug;90(5):412-417 [FREE Full text] [doi: 10.1111/j.1755-3768.2010.02028.x] [Medline: 21054818]
41. Saleh GM, Lamparter J, Sullivan PM, O'Sullivan F, Hussain B, Athanasiadis I, et al. The international forum of ophthalmic simulation: developing a virtual reality training curriculum for ophthalmology. Br J Ophthalmol 2013 Jun 26;97(6):789-792. [doi: 10.1136/bjophthalmol-2012-302764] [Medline: 23532612]
42. Gonzalez-Gonzalez LA, Payal AR, Gonzalez-Monroy JE, Daly MK. Ophthalmic surgical simulation in training dexterity in dominant and nondominant hands: results from a pilot study. J Surg Educ 2016;73(4):699-708. [doi: 10.1016/j.jsurg.2016.01.014] [Medline: 27017524]
43. Bozkurt Oflaz A, Ekinci Köktekir B, Okudan S. Does cataract surgery simulation correlate with real-life experience? Turk J Ophthalmol 2018 Jun;48(3):122-126 [FREE Full text] [doi: 10.4274/tjo.10586] [Medline: 29988849]
44. Ropelato S, Menozzi M, Michel D, Siegrist M. Augmented reality microsurgery: a tool for training micromanipulations in ophthalmic surgery using augmented reality. Simul Healthc 2020 Apr;15(2):122-127. [doi: 10.1097/SIH.0000000000000413] [Medline: 32044852]
45. Jonas JB, Rabethge S, Bender H. Computer-assisted training system for pars plana vitrectomy. Acta Ophthalmol Scand 2003 Dec;81(6):600-604 [FREE Full text] [doi: 10.1046/j.1395-3907.2003.0078.x] [Medline: 14641261]
46. Feldman BH, Ake JM, Geist CE. Virtual reality simulation. Ophthalmology 2007 Apr;114(4):828.e1-828.e4. [doi: 10.1016/j.ophtha.2006.10.016] [Medline: 17398334]
47. Feudner EM, Engel C, Neuhann IM, Petermeier K, Bartz-Schmidt K, Szurman P. Virtual reality training improves wet-lab performance of capsulorhexis: results of a randomized, controlled study. Graefes Arch Clin Exp Ophthalmol 2009 Jul;247(7):955-963. [doi: 10.1007/s00417-008-1029-7] [Medline: 19172289]
48. Peugnet F, Dubois P, Rouland JF. Virtual reality versus conventional training in retinal photocoagulation: a first clinical assessment. Comput Aided Surg 1998;3(1):20-26. [doi: 10.1002/(SICI)1097-0150(1998)3:1<20::AID-IGS3>3.0.CO;2-N] [Medline: 9699075]
49. Daly MK, Gonzalez E, Siracuse-Lee D, Legutko PA. Efficacy of surgical simulator training versus traditional wet-lab training on operating room performance of ophthalmology residents during the capsulorhexis in cataract surgery. J Cataract Refract Surg 2013 Nov;39(11):1734-1741. [doi: 10.1016/j.jcrs.2013.05.044] [Medline: 24160383]
50. Deuchler S, Wagner C, Singh P, Müller M, Al-Dwairi R, Benjilali R, et al. Clinical efficacy of simulated vitreoretinal surgery to prepare surgeons for the upcoming intervention in the operating room. PLoS One 2016;11(3):e0150690 [FREE Full text] [doi: 10.1371/journal.pone.0150690] [Medline: 26964040]
51. Alwadani F, Morsi MS. PixEye virtual reality training has the potential of enhancing proficiency of laser trabeculoplasty performed by medical students: a pilot study. Middle East Afr J Ophthalmol 2012 Jan;19(1):120-122 [FREE Full text] [doi: 10.4103/0974-9233.92127] [Medline: 22346126]
52. Sikder S, Luo J, Banerjee PP, Luciano C, Kania P, Song JC, et al. The use of a virtual reality surgical simulator for cataract surgical skill assessment with 6 months of intervening operating room experience. Clin Ophthalmol 2015;9:141-149 [FREE Full text] [doi: 10.2147/OPTH.S69970] [Medline: 25653496]
53. Saleh GM, Theodoraki K, Gillan S, Sullivan P, O'Sullivan F, Hussain B, et al. The development of a virtual reality training programme for ophthalmology: repeatability and reproducibility (part of the International Forum for Ophthalmic Simulation Studies). Eye (Lond) 2013 Nov;27(11):1269-1274 [FREE Full text] [doi: 10.1038/eye.2013.166] [Medline: 23970027]

54. Rossi JV, Verma D, Fujii GY, Lakhanpal RR, Wu SL, Humayun MS, et al. Virtual vitreoretinal surgical simulator as a training tool. Retina 2004 Apr;24(2):231-236. [doi: 10.1097/00006982-200404000-00007] [Medline: 15097883]
55. Mahr MA, Hodge DO. Construct validity of anterior segment anti-tremor and forceps surgical simulator training modules: attending versus resident surgeon performance. J Cataract Refract Surg 2008 Jun;34(6):980-985. [doi: 10.1016/j.jcrs.2008.02.015] [Medline: 18499005]
56. Solverson DJ, Mazzoli RA, Raymond WR, Nelson ML, Hansen EA, Torres MF, et al. Virtual reality simulation in acquiring and differentiating basic ophthalmic microsurgical skills. Simul Healthc 2009;4(2):98-103. [doi: 10.1097/SIH.0b013e318195419e] [Medline: 19444047]
57. Privett B, Greenlee E, Rogers G, Oetting TA. Construct validity of a surgical simulator as a valid model for capsulorhexis training. J Cataract Refract Surg 2010 Nov;36(11):1835-1838. [doi: 10.1016/j.jcrs.2010.05.020] [Medline: 21029889]
58. Nathoo N, Ng M, Ramstead CL, Johnson MC. Comparing performance of junior and senior ophthalmology residents on an intraocular surgical simulator. Can J Ophthalmol 2011 Feb;46(1):87-88. [doi: 10.3129/i10-065] [Medline: 21283165]
59. Le TDB, Adatia FA, Lam WC. Virtual reality ophthalmic surgical simulation as a feasible training and assessment tool: results of a multicentre study. Can J Ophthalmol 2011 Feb;46(1):56-60. [doi: 10.3129/i10-051] [Medline: 21283159]
60. Cissé C, Angioi K, Luc A, Berrod JP, Conart JB. EYESI surgical simulator: validity evidence of the vitreoretinal modules. Acta Ophthalmol 2019 Mar;97(2):e277-e282 [FREE Full text] [doi: 10.1111/aos.13910] [Medline: 30168257]
61. Spiteri AV, Aggarwal R, Kersey TL, Sira M, Benjamin L, Darzi AW, et al. Development of a virtual reality training curriculum for phacoemulsification surgery. Eye (Lond) 2014 Jan;28(1):78-84 [FREE Full text] [doi: 10.1038/eye.2013.211] [Medline: 24071776]
62. Lam CK, Sundaraj K, Sulaiman MN, Qamarruddin FA. Virtual phacoemulsification surgical simulation using visual guidance and performance parameters as a feasible proficiency assessment tool. BMC Ophthalmol 2016 Jun 14;16:88 [FREE Full text] [doi: 10.1186/s12886-016-0269-2] [Medline: 27296449]
63. Selvander M, Asman P. Ready for OR or not? human reader supplements Eyesi scoring in cataract surgical skills assessment. Clin Ophthalmol 2013;7:1973-1977 [FREE Full text] [doi: 10.2147/OPTH.S48374] [Medline: 24124350]
64. Thomsen ASS, Kiilgaard JF, Kjaerbo H, la Cour M, Konge L. Simulation-based certification for cataract surgery. Acta Ophthalmol 2015 Aug;93(5):416-421 [FREE Full text] [doi: 10.1111/aos.12691] [Medline: 25722080]
65. Thomsen ASS, Smith P, Subhi Y, Cour ML, Tang L, Saleh GM, et al. High correlation between performance on a virtual-reality simulator and real-life cataract surgery. Acta Ophthalmol 2017 May;95(3):307-311 [FREE Full text] [doi: 10.1111/aos.13275] [Medline: 27679989]
66. Jacobsen MF, Konge L, Bach-Holm D, la Cour M, Holm L, Højgaard-Olsen K, et al. Correlation of virtual reality performance with real-life cataract surgery performance. J Cataract Refract Surg 2019 Sep;45(9):1246-1251. [doi: 10.1016/j.jcrs.2019.04.007] [Medline: 31371151]
67. Vergmann AS, Vestergaard AH, Grauslund J. Virtual vitreoretinal surgery: validation of a training programme. Acta Ophthalmol 2017 Feb;95(1):60-65 [FREE Full text] [doi: 10.1111/aos.13209] [Medline: 27535480]
68. Selvander M, Asman P. Cataract surgeons outperform medical students in Eyesi virtual reality cataract surgery: evidence for construct validity. Acta Ophthalmol 2013 Aug;91(5):469-474 [FREE Full text] [doi: 10.1111/j.1755-3768.2012.02440.x] [Medline: 22676143]
69. Wu DJ, Greenberg PB. A self-directed preclinical course in ophthalmic surgery. J Surg Educ 2016;73(3):370-374. [doi: 10.1016/j.jsurg.2015.11.005] [Medline: 26705060]
70. Yong JJ, Migliori ME, Greenberg PB. A novel preclinical course in ophthalmology and ophthalmic virtual surgery. Med Health R I 2012 Nov;95(11):345-348. [Medline: 23477278]
71. Liang S, Banerjee PP, Edward DP. A high performance graphic and haptic curvilinear capsulorrhexis simulation system. Conf Proc IEEE Eng Med Biol Soc 2009;2009:5092-5095. [doi: 10.1109/IEMBS.2009.5332727] [Medline: 19963878]
72. Laurell C, Söderberg P, Nordh L, Skarman E, Nordqvist P. Computer-simulated phacoemulsification. Ophthalmology 2004 Apr;111(4):693-698. [doi: 10.1016/j.ophtha.2003.06.023] [Medline: 15051200]
73. Ng DS, Sun Z, Young AL, Ko ST, Lok JK, Lai TY, et al. Impact of virtual reality simulation on learning barriers of phacoemulsification perceived by residents. Clin Ophthalmol 2018;12:885-893 [FREE Full text] [doi: 10.2147/OPTH.S140411] [Medline: 29785084]
74. Leitritz MA, Ziemssen F, Suesskind D, Partsch M, Voykov B, Bartz-Schmidt KU, et al. Critical evaluation of the usability of augmented reality ophthalmoscopy for the training of inexperienced examiners. Retina 2014 Apr;34(4):785-791. [doi: 10.1097/IAE.0b013e3182a2e75d] [Medline: 24670999]
75. Rai AS, Rai AS, Mavrikakis E, Lam WC. Teaching binocular indirect ophthalmoscopy to novice residents using an augmented reality simulator. Can J Ophthalmol 2017 Oct;52(5):430-434. [doi: 10.1016/j.jcjo.2017.02.015] [Medline: 28985799]
76. Chou J, Kosowsky T, Payal AR, Gonzalez Gonzalez LA, Daly MK. Construct and face validity of the eyesi indirect ophthalmoscope simulator. Retina 2017 Oct;37(10):1967-1976. [doi: 10.1097/IAE.0000000000001438] [Medline: 28045850]
77. Borgersen NJ, Skou Thomsen AS, Konge L, Sørensen TL, Subhi Y. Virtual reality-based proficiency test in direct ophthalmoscopy. Acta Ophthalmol 2018 Mar;96(2):e259-e261 [FREE Full text] [doi: 10.1111/aos.13546] [Medline: 28834332]

78. Nguyen M, Quevedo-Uribe A, Kapralos B, Jenkin M, Kanev K, Jaimes N. An experimental training support framework for eye fundus examination skill development. Comput Methods Biomech Biomed Eng Imaging Vis 2017 Oct 12;7(1):26-36. [doi: 10.1080/21681163.2017.1376708]
79. Wilson AS, O'Connor J, Taylor L, Carruthers D. A 3D virtual reality ophthalmoscopy trainer. Clin Teach 2017 Dec;14(6):427-431. [doi: 10.1111/tct.12646] [Medline: 28401705]
80. Wei L, Najdovski Z, Nahavandi S, Weisinger H. Towards a haptically enabled optometry training simulator. Netw Model Anal Health Inform Bioinforma 2014 May 1;3(1):1-8. [doi: 10.1007/s13721-014-0060-3]
81. Wei L, Najdovski Z, Wael A, Nahavandi S, Weisinger H. Augmented optometry training simulator with multi-point haptics. In: Proceedings of the 2012 IEEE International Conference on Systems, Man, and Cybernetics. 2012 Presented at: IEEE International Conference on Systems, Man, and Cybernetics; October 14-17; Seoul, South Korea p. 2991-2997. [doi: 10.1109/icsmc.2012.6378250]
82. Maloca PM, de Carvalho JER, Heeren T, Hasler PW, Mushtaq F, Mon-Williams M, et al. High-performance virtual reality volume rendering of original optical coherence tomography point-cloud data enhanced with real-time ray casting. Transl Vis Sci Technol 2018 Jul;7(4):2 [FREE Full text] [doi: 10.1167/tvst.7.4.2] [Medline: 30002949]
83. Berger JW, Madjarov B. Augmented reality fundus biomicroscopy: a working clinical prototype. Arch Ophthalmol 2001 Dec;119(12):1815-1818. [doi: 10.1001/archopht.119.12.1815] [Medline: 11735793]
84. Goh RLZ, Kong YXG, McAlinden C, Liu J, Crowston JG, Skalicky SE. Objective assessment of activity limitation in glaucoma with smartphone virtual reality goggles: a pilot study. Transl Vis Sci Technol 2018 Jan;7(1):10 [FREE Full text] [doi: 10.1167/tvst.7.1.10] [Medline: 29372112]
85. Ungewiss J, Kübler T, Sippel K, Aehling K, Heister M, Rosenstiel W, Simulator/On-road Study Group. Agreement of driving simulator and on-road driving performance in patients with binocular visual field loss. Graefes Arch Clin Exp Ophthalmol 2018 Dec;256(12):2429-2435. [doi: 10.1007/s00417-018-4148-9] [Medline: 30251198]
86. Jones PR, Somoskeöy T, Chow-Wing-Bom H, Crabb DP. Seeing other perspectives: evaluating the use of virtual and augmented reality to simulate visual impairments (OpenVisSim). NPJ Digit Med 2020;3:32 [FREE Full text] [doi: 10.1038/s41746-020-0242-6] [Medline: 32195367]
87. Asfaw DS, Jones PR, Mönter VM, Smith ND, Crabb DP. Does glaucoma alter eye movements when viewing images of natural scenes? a between-eye study. Invest Ophthalmol Vis Sci 2018 Jul 02;59(8):3189-3198. [doi: 10.1167/iovs.18-23779] [Medline: 29971443]
88. Dive S, Rouland JF, Lenoble Q, Szaffarczyk S, McKendrick AM, Boucart M. Impact of peripheral field loss on the execution of natural actions: a study with glaucomatous patients and normally sighted people. J Glaucoma 2016 Oct;25(10):e889-e896. [doi: 10.1097/IJG.0000000000000402] [Medline: 26950583]
89. Lee SS, Black AA, Wood JM. Effect of glaucoma on eye movement patterns and laboratory-based hazard detection ability. PLoS One 2017;12(6):e0178876 [FREE Full text] [doi: 10.1371/journal.pone.0178876] [Medline: 28570621]
90. Qian Z, Wang H, Fan H, Lin D, Li W. Three-dimensional digital visualization of phacoemulsification and intraocular lens implantation. Indian J Ophthalmol 2019 Mar;67(3):341-343. [doi: 10.4103/ijo.IJO_1012_18] [Medline: 30777950]
91. Talcott KE, Adam MK, Sioufi K, Aderman CM, Ali FS, Mellen PL, et al. Comparison of a three-dimensional heads-up display surgical platform with a standard operating microscope for macular surgery. Ophthalmol Retina 2019 Mar;3(3):244-251. [doi: 10.1016/j.oret.2018.10.016] [Medline: 31014702]
92. Romano MR, Cennamo G, Comune C, Cennamo M, Ferrara M, Rombetto L, et al. Evaluation of 3D heads-up vitrectomy: outcomes of psychometric skills testing and surgeon satisfaction. Eye (Lond) 2018 Jun;32(6):1093-1098 [FREE Full text] [doi: 10.1038/s41433-018-0027-1] [Medline: 29445116]
93. Kumar A, Hasan N, Kakkar P, Mutha V, Karthikeya R, Sundar D, et al. Comparison of clinical outcomes between "heads-up" 3D viewing system and conventional microscope in macular hole surgeries: a pilot study. Indian J Ophthalmol 2018 Dec;66(12):1816-1819 [FREE Full text] [doi: 10.4103/ijo.IJO_59_18] [Medline: 30451186]
94. Zhang Z, Wang L, Wei Y, Fang D, Fan S, Zhang S. The preliminary experiences with three-dimensional heads-up display viewing system for vitreoretinal surgery under various status. Curr Eye Res 2019 Jan;44(1):102-109. [doi: 10.1080/02713683.2018.1526305] [Medline: 30265818]
95. Matsumoto CS, Shibuya M, Makita J, Shoji T, Ohno H, Shinoda K, et al. Heads-up 3d surgery under low light intensity conditions: new high-sensitivity hd camera for ophthalmological microscopes. J Ophthalmol 2019;2019:5013463 [FREE Full text] [doi: 10.1155/2019/5013463] [Medline: 31885885]
96. Rizzo S, Abbruzzese G, Savastano A, Giansanti F, Caporossi T, Barca F, et al. 3D surgical viewing system in ophthalmology: perceptions of the surgical team. Retina 2018 Apr;38(4):857-861. [doi: 10.1097/IAE.0000000000002018] [Medline: 29283980]
97. Martínez-Toldos JJ, Fernández-Martínez C, Navarro-Navarro A. Experience using a 3d head-mounted display system in ophthalmic surgery. Retina 2017 Jul;37(7):1419-1421. [doi: 10.1097/IAE.0000000000001664] [Medline: 28492431]
98. Eckardt C, Paulo EB. Heads-up surgery for vitreoretinal procedures: an experimental and clinical study. Retina 2016 Jan;36(1):137-147. [doi: 10.1097/IAE.0000000000000689] [Medline: 26200516]
99. Kunikata H, Abe T, Nakazawa T. Heads-up macular surgery with a 27-gauge microincision vitrectomy system and minimal illumination. Case Rep Ophthalmol 2016;7(3):265-269 [FREE Full text] [doi: 10.1159/000452993] [Medline: 28101044]

100. Coppola M, La Spina C, Rabiolo A, Querques G, Bandello F. Heads-up 3D vision system for retinal detachment surgery. Int J Retina Vitreous 2017;3:46 [FREE Full text] [doi: 10.1186/s40942-017-0099-2] [Medline: 29188074]
101. Kita M, Mori Y, Hama S. Hybrid wide-angle viewing-endoscopic vitrectomy using a 3D visualization system. Clin Ophthalmol 2018;12:313-317 [FREE Full text] [doi: 10.2147/OPTH.S156497] [Medline: 29491704]
102. Kita M, Kusaka M, Yamada H, Hama S. Three-dimensional ocular endoscope system for vitrectomy. Clin Ophthalmol 2019;13:1641-1643 [FREE Full text] [doi: 10.2147/OPTH.S221197] [Medline: 31695313]
103. Kita M, Kusaka M, Yamada H, Hama S. Updated chandelier illumination-assisted scleral buckling using 3D visualization system. Clin Ophthalmol 2019;13:1743-1748 [FREE Full text] [doi: 10.2147/OPTH.S218975] [Medline: 31564825]
104. Agranat JS, Miller JB, Douglas VP, Douglas KAA, Marmalidou A, Cunningham MA, et al. The scope of three-dimensional digital visualization systems in vitreoretinal surgery. Clin Ophthalmol 2019;13:2093-2096 [FREE Full text] [doi: 10.2147/OPTH.S213834] [Medline: 31749606]
105. Zhang T, Tang W, Xu G. Comparative analysis of three-dimensional heads-up vitrectomy and traditional microscopic vitrectomy for vitreoretinal diseases. Curr Eye Res 2019 Oct;44(10):1080-1086. [doi: 10.1080/02713683.2019.1612443] [Medline: 31021174]
106. Palácios RM, de Carvalho ACM, Maia M, Caiado RR, Camilo DAG, Farah ME. An experimental and clinical study on the initial experiences of Brazilian vitreoretinal surgeons with heads-up surgery. Graefes Arch Clin Exp Ophthalmol 2019 Mar;257(3):473-483. [doi: 10.1007/s00417-019-04246-w] [Medline: 30645695]
107. Ali MJ, Naik MN. First intraoperative experience with three-dimensional (3D) high-definition (HD) nasal endoscopy for lacrimal surgeries. Eur Arch Otorhinolaryngol 2017 May;274(5):2161-2164. [doi: 10.1007/s00405-017-4466-2] [Medline: 28154929]
108. Ehlers JP, Uchida A, Srivastava SK. The integrative surgical theater: combining intraoperative optical coherence tomography and 3d digital visualization for vitreoretinal surgery in the DISCOVER study. Retina 2018 Sep;38 Suppl 1:S88-S96 [FREE Full text] [doi: 10.1097/IAE.0000000000001999] [Medline: 29256988]
109. Freeman WR, Chen KC, Ho J, Chao DL, Ferreyra HA, Tripathi AB, et al. Resolution, depth of field, and physician satisfaction during digitally assisted vitreoretinal surgery. Retina 2019 Sep;39(9):1768-1771 [FREE Full text] [doi: 10.1097/IAE.0000000000002236] [Medline: 29965938]
110. Bhadri PR, Rowley AP, Khurana RN, Deboer CM, Kerns RM, Chong LP, et al. Evaluation of a stereoscopic camera-based three-dimensional viewing workstation for ophthalmic surgery. Am J Ophthalmol 2007 May;143(5):891-892. [doi: 10.1016/j.ajo.2006.12.032] [Medline: 17452184]
111. Lee H, Kim S. Effectiveness of binocularity-stimulating treatment in children with residual amblyopia following occlusion. BMC Ophthalmol 2018 Sep 20;18(1):253. [doi: 10.1186/s12886-018-0922-z] [Medline: 30236086]
112. Sayed AM, Abdel-Mottaleb M, Kashem R, Roongpoovapatr V, Elsawy A, Abdel-Mottaleb M, et al. Expansion of peripheral visual field with novel virtual reality digital spectacles. Am J Ophthalmol 2020 Feb;210:125-135. [doi: 10.1016/j.ajo.2019.10.006] [Medline: 31626763]
113. Sánchez J, Tadres A. Augmented reality application for the navigation of people who are blind. Int J Disabil Hum Dev 2011 Jan 1;10(1):75-79. [doi: 10.1515/ijdhd.2011.015]
114. Tobler-Ammann BC, Surer E, de Bruin ED, Rabuffetti M, Borghese NA, Mainetti R, et al. Exergames encouraging exploration of hemineglected space in stroke patients with visuospatial neglect: a feasibility study. JMIR Serious Games 2017 Aug 25;5(3):e17 [FREE Full text] [doi: 10.2196/games.7923] [Medline: 28842388]
115. Melillo P, Riccio D, Di Perna L, Sanniti Di Baja G, De Nino M, Rossi S, et al. Wearable improved vision system for color vision deficiency correction. IEEE J Transl Eng Health Med 2017;5:3800107 [FREE Full text] [doi: 10.1109/JTEHM.2017.2679746] [Medline: 28507827]
116. Fitts P, Posner M. Human Performance. Oxford, England: Brooks/Cole; 1967.
117. Dreyfus SE. The five-stage model of adult skill acquisition. Bull Sci Technol Soc 2016 Jul 24;24(3):177-181. [doi: 10.1177/0270467604264992]
118. Nicholls D, Sweet L, Muller A, Hyett J. Teaching psychomotor skills in the twenty-first century: revisiting and reviewing instructional approaches through the lens of contemporary literature. Med Teach 2016 Oct;38(10):1056-1063. [doi: 10.3109/0142159X.2016.1150984] [Medline: 27023405]
119. Liu JJ, Witkin AJ, Adhi M, Grulkowski I, Kraus MF, Dhalla A, et al. Enhanced vitreous imaging in healthy eyes using swept source optical coherence tomography. PLoS One 2014;9(7):e102950 [FREE Full text] [doi: 10.1371/journal.pone.0102950] [Medline: 25036044]
120. Maloca P, Gyger C, Hasler PW. A pilot study to image the vascular network of small melanocytic choroidal tumors with speckle noise-free 1050-nm swept source optical coherence tomography (OCT choroidal angiography). Graefes Arch Clin Exp Ophthalmol 2016 Jun;254(6):1201-1210. [doi: 10.1007/s00417-015-3259-9] [Medline: 26847040]
121. Waddingham PE, Butler TKH, Cobb SV, Moody ADR, Comaish IF, Haworth SM, et al. Preliminary results from the use of the novel Interactive binocular treatment (I-BiT) system, in the treatment of strabismic and anisometropic amblyopia. Eye (Lond) 2006 Mar;20(3):375-378. [doi: 10.1038/sj.eye.6701883] [Medline: 15832181]
122. Cleary M, Moody AD, Buchanan A, Stewart H, Dutton GN. Assessment of a computer-based treatment for older amblyopes: the Glasgow pilot study. Eye (Lond) 2009 Jan;23(1):124-131. [doi: 10.1038/sj.eye.6702977] [Medline: 17932508]

123. Birch EE, Li SL, Jost RM, Morale SE, De La Cruz A, Stager D, et al. Binocular iPad treatment for amblyopia in preschool children. J AAPOS 2015 Feb;19(1):6-11 [FREE Full text] [doi: 10.1016/j.jaapos.2014.09.009] [Medline: 25727578]
124. Li SL, Jost RM, Morale SE, Stager DR, Dao L, Stager D, et al. A binocular iPad treatment for amblyopic children. Eye (Lond) 2014 Oct;28(10):1246-1253 [FREE Full text] [doi: 10.1038/eye.2014.165] [Medline: 25060850]
125. Wee SW, Moon NJ, Lee WK, Jeon S. Ophthalmological factors influencing visual asthenopia as a result of viewing 3D displays. Br J Ophthalmol 2012 Nov;96(11):1391-1394. [doi: 10.1136/bjophthalmol-2012-301690] [Medline: 22952400]
126. Yum HR, Park SH, Kang H, Shin SY. Changes in ocular factors according to depth variation and viewer age after watching a three-dimensional display. Br J Ophthalmol 2014 May;98(5):684-690. [doi: 10.1136/bjophthalmol-2013-304244] [Medline: 24457372]
127. Yum HR, Park SH, Kang H, Shin SY. Changes in ocular parameters depending on the motion-in-depth of a three-dimensional display. Br J Ophthalmol 2015 Dec;99(12):1706-1712. [doi: 10.1136/bjophthalmol-2014-306297] [Medline: 26024674]
128. Wee SW, Moon NJ. Clinical evaluation of accommodation and ocular surface stability relevant to visual asthenopia with 3D displays. BMC Ophthalmol 2014 Mar 11;14:29 [FREE Full text] [doi: 10.1186/1471-2415-14-29] [Medline: 24612686]
129. Morse SE, Jiang BC. Oculomotor function after virtual reality use differentiates symptomatic from asymptomatic individuals. Optom Vis Sci 1999 Sep;76(9):637-642. [doi: 10.1097/00006324-199909000-00021] [Medline: 10498005]
130. Ha S, Na K, Kweon I, Suh Y, Kim S. Effects of head-mounted display on the oculomotor system and refractive error in normal adolescents. J Pediatr Ophthalmol Strabismus 2016 Jul 01;53(4):238-245. [doi: 10.3928/01913913-20160511-01] [Medline: 27281827]
131. Tychsen L, Foeller P. Effects of immersive virtual reality headset viewing on young children: visuomotor function, postural stability, and motion sickness. Am J Ophthalmol 2020 Jan;209:151-159. [doi: 10.1016/j.ajo.2019.07.020] [Medline: 31377280]
132. Farmani Y, Teather RJ. Evaluating discrete viewpoint control to reduce cybersickness in virtual reality. Virtual Reality 2020 Jan 13;24(4):645-664. [doi: 10.1007/s10055-020-00425-x]
133. Young BK, Greenberg PB. Is virtual reality training for resident cataract surgeons cost effective? Graefes Arch Clin Exp Ophthalmol 2013 Mar 28;251(9):2295-2296. [doi: 10.1007/s00417-013-2317-4]
134. Lowry EA, Porco TC, Naseri A. Cost analysis of virtual-reality phacoemulsification simulation in ophthalmology training programs. J Cataract Refract Surg 2013 Oct;39(10):1616-1617. [doi: 10.1016/j.jcrs.2013.08.015] [Medline: 24075164]

Abbreviations

DSpecs: digital spectacles
GRASIS: Global Rating Assessment of Skills in Intraocular Surgery
OCEBM: Oxford Centre for Evidence-Based Medicine
OSACSS: Objective Structured Assessment of Cataract Surgical Skill Score
OSATS: Objective Structured Assessment of Technical Surgical Skills
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-analyses

Edited by G Eysenbach; submitted 08.09.20; peer-reviewed by D Gunasekeran, E Surer; comments to author 18.11.20; revised version received 19.12.20; accepted 06.04.21; published 19.08.21.

Please cite as:
Ong CW, Tan MCJ, Lam M, Koh VTC
Applications of Extended Reality in Ophthalmology: Systematic Review
J Med Internet Res 2021;23(8):e24152
URL: https://www.jmir.org/2021/8/e24152
doi:10.2196/24152
PMID:34420929

©Chee Wui Ong, Marcus Chun Jin Tan, Michael Lam, Victor Teck Chang Koh. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Variation of Daily Care Demand in Swiss General Hospitals: Longitudinal Study on Capacity Utilization, Patient Turnover and Clinical Complexity Levels

Narayan Sharma[1], MSc; René Schwendimann[1,2], PhD; Olga Endrich[3], MD; Dietmar Ausserhofer[1,4], PhD; Michael Simon[1,5], PhD

[1]Institute of Nursing Science, Department of Public Health, Faculty of Medicine, University of Basel, Basel, Switzerland
[2]Patient Safety Office, University Hospital Basel, Basel, Switzerland
[3]Directorate of Medicine, Inselspital University Hospital Bern, Bern, Switzerland
[4]College of Health-Care Professions Claudiana, Bozen, Italy
[5]Nursing Research Unit, Inselspital University Hospital Bern, Bern, Switzerland

Corresponding Author:
Michael Simon, PhD
Institute of Nursing Science
Department Public Health
Faculty of Medicine, University of Basel
Bernoullistrasse 28
Basel, 4056
Switzerland
Phone: 41 61 207 09 12
Fax: 41 61 207 09 55
Email: m.simon@unibas.ch

Abstract

Background: Variations in hospitals' care demand relies not only on the patient volume but also on the disease severity. Understanding both daily severity and patient volume in hospitals could help to identify hospital pressure zones to improve hospital-capacity planning and policy-making.

Objective: This longitudinal study explored daily care demand dynamics in Swiss general hospitals for 3 measures: (1) capacity utilization, (2) patient turnover, and (3) patient clinical complexity level.

Methods: A retrospective population-based analysis was conducted with 1 year of routine data of 1.2 million inpatients from 102 Swiss general hospitals. Capacity utilization was measured as a percentage of the daily maximum number of inpatients. Patient turnover was measured as a percentage of the daily sum of admissions and discharges per hospital. Patient clinical complexity level was measured as the average daily patient disease severity per hospital from the clinical complexity algorithm.

Results: There was a pronounced variability of care demand in Swiss general hospitals. Among hospitals, the average daily capacity utilization ranged from 57.8% (95% CI 57.3-58.4) to 87.7% (95% CI 87.3-88.0), patient turnover ranged from 22.5% (95% CI 22.1-22.8) to 34.5% (95% CI 34.3-34.7), and the mean patient clinical complexity level ranged from 1.26 (95% CI 1.25-1.27) to 2.06 (95% CI 2.05-2.07). Moreover, both within and between hospitals, all 3 measures varied distinctly between days of the year, between days of the week, between weekdays and weekends, and between seasons.

Conclusions: While admissions and discharges drive capacity utilization and patient turnover variation, disease severity of each patient drives patient clinical complexity level. Monitoring—and, if possible, anticipating—daily care demand fluctuations is key to managing hospital pressure zones. This study provides a pathway for identifying patients' daily exposure to strained hospital systems for a time-varying causal model.

(J Med Internet Res 2021;23(8):e27163) doi:10.2196/27163

KEYWORDS

inpatient population; routine data; general hospitals; capacity utilization; clinical complexity; patient data; hospital system; complexity algorithm

Introduction

Hospitals are constantly challenged by changing patient care demands. If this outweighs available resources, it can affect the quality of care and patient safety [1]. Demand factors include daily patient volume, turnover, and clinical complexity of patients requiring diagnoses and therapies [1,2]. Responding to variations in any of these factors, hospitals adjust their resource supplies (eg, by changing shift-level staffing or resources for each day, between workdays and weekends, for different seasons, and throughout the year) [3,4].

Capacity utilization, which is based on the number of beds occupied vs those available [5], offers one perspective to view hospital care demand [2,6,7]. Over recent decades, most health care systems' capacity utilization has increased, while total numbers of available beds have decreased [3,8]. This trend mainly reflects policies to reduce health care costs and to increase efficiency (eg, by the use of diagnostic-related groups [DRGs]) [8,9].

If capacity utilization is too high (eg, above 80% or 85%), it might overburden health care systems and their workforces [10,11], possibly leading to adverse patient outcomes such as infections or even death [5,12,13]. Capacity utilization is high—exceeding 90%—in Canada, Israel, and Ireland, followed by the United Kingdom, Norway, and Switzerland, all of which report figures above 80% [14]. Within a country, capacity utilization also varies between hospital types, geographic regions, and populations served [6,15]. In Switzerland, the most recent annual capacity utilization figures for acute care hospitals were between 70% and 82% [16,17].

As noted above, care demand also relies on patient turnover and patient complexity [18]. Patient turnover refers to the admission and the discharge or transfer of patients between units or hospitals [19], requiring resources [20]. "Census variability," "churn," or "environmental turbulence" cover the same or similar concepts [21]. Patient complexity refers to the severity or complexity of each patient's clinical needs. For instance, patients admitted to the intensive care unit generally require more resources than those in a general ward, representing a resource-intensive caseload.

Disease severity is commonly measured via the Charlson-Elixhauser comorbidity or case-mix index; however, this does not include all relevant clinical conditions or morbidities, and its interpretation is commonly influenced by reimbursement policies or the cost of medication and treatment [22-24]. One alternative measure is the patient clinical complexity level. As part of the German DRG system, the patient clinical complexity level reflects upon not only complications and comorbidities but also their levels of clinical severity on a 5-point scale (ie, 0=no, 1=mild, 2=moderate, 3=severe, and 4=very severe clinical complexity) [25]. A complex algorithm, depending on primary and secondary diagnoses and estimated severities, allows determination of the cumulative effect of the diagnoses per treatment episode [22,26,27]. A higher patient clinical complexity level indicates a more complex and resource-intensive caseload.

As capacity utilization, patient turnover, and patient clinical complexity level all offer necessary perspectives on hospital care demand, all 3 are vital for optimal resource allocation [28]. All 3 are also connected; for instance, complex patients usually stay longer in hospitals, increasing capacity utilization. Furthermore, with each transfer or referral to another hospital, additional load is created, as patients need to be assessed at admission or be prepared for discharge [21]. Understanding these factors' daily variation is a vital step toward optimizing health care structures and processes [29]. Regarding the daily variability of care demand, analysis of long-term data can help to anticipate when staffs or supplies will be depleted or strained, thereby, indicating when and where to allocate resources. Particularly, the traditional measures of capacity utilization (eg, midnight count [29,30]) and patient turnover (eg, the inverse of the length of hospital stay [31,32]) may not convey the dynamic nature of actual daytime hospital care demands [33]; certainly, neither incorporates patient complexity and severity.

Additionally, the valid, highly granular longitudinal (daily or weekly) measurement of care demand offers a precise and in-depth view of how care needs fluctuate and evolve. As capacity utilization, patient turnover, and patient clinical complexity level are time-sensitive variables, the 3 together offer great potential to accurately represent daily care demand dynamics. Such information should enable health care managers to anticipate capacity needs to accommodate patients during a typical weekday, weekend, or seasonal peak [3].

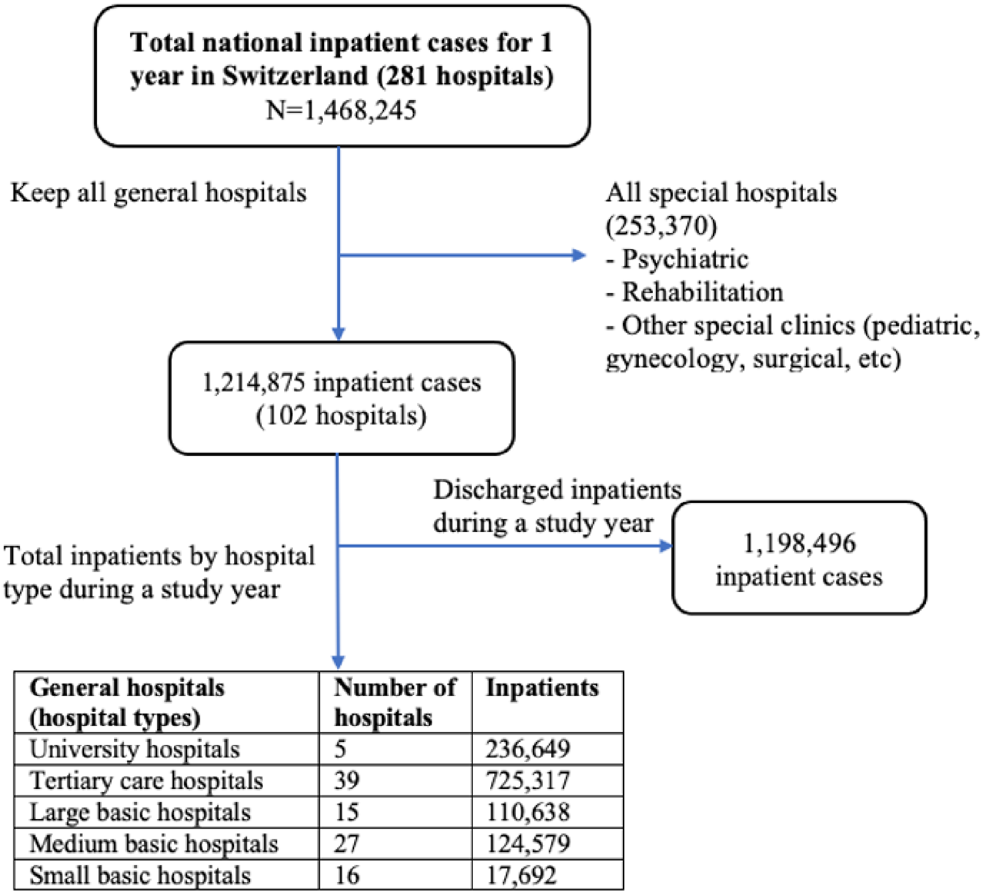
Therefore, this study aimed to describe the daily care demand in general hospitals from a longitudinal perspective, specifically, the daily peaks and variations during weeks (ie, weekends vs weekdays), as well as seasons. This study describes the daily variability of (1) capacity utilization, (2) patient turnover, and (3) patient clinical complexity levels of Swiss general hospitals' inpatient populations.

Methods

Study Design, Setting, and Population

This is a retrospective, population-based analysis using 1 year of hospital data extracted from the 6-year (2012-2017) dataset obtained from the Swiss Federal Statistics Office (FSO). Based on a data protection contract (as stipulated by article 22 of the Swiss Federal Act on Data Protection), the FSO provided anonymized data from all Swiss hospital inpatients hospitalized over the study period. The data covered general as well as specialized care facilities such as pediatric, gynecological, psychiatric, and rehabilitative hospitals [34]. The FSO divides general hospitals into 5 classifications: university hospitals, tertiary care hospitals, large basic hospitals, medium basic hospitals, and small basic hospitals. Classification is based on the number of cases treated per year and a weighted sum of service points (based on a combination of the number of hospital units and the levels of care delivered) assigned by the Swiss Medical Association [17,34]. For instance, based on the Swiss Medical Association classification, a university hospital requires the weighted sum of service points to be >100 units and >30,000 cases per year [34]. For this study, we included a 1-year patient population dataset from general hospitals to limit interhospital heterogeneity. Due to the Swiss Data Protection Act's stipulations regarding patient anonymity, we were unable to trace patients across years. An overview of the process of selecting inpatient cases for analysis is included in the flow diagram in Figure 1.

Figure 1. Flow diagram of inpatient cases for the analysis.


Total national inpatient cases for 1 year in Switzerland (281 hospitals)
N=1,468,245
Keep all general hospitals
All special hospitals (253,370)
- Psychiatric
- Rehabilitation
- Other special clinics (pediatric, gynecology, surgical, etc)
1,214,875 inpatient cases (102 hospitals)
Discharged inpatients during a study year
1,198,496 inpatient cases
Total inpatients by hospital type during a study year

General hospitals (hospital types)	Number of hospitals	Inpatients
University hospitals	5	236,649
Tertiary care hospitals	39	725,317
Large basic hospitals	15	110,638
Medium basic hospitals	27	124,579
Small basic hospitals	16	17,692

Dataset and Study Variables

We extracted 1-year datasets, including (1) all inpatients discharged during each study year, regardless of admission date, and (2) all inpatients admitted during each study year, regardless of discharge date. From the full FSO hospital dataset, we also extracted data for all relevant routine administrative and clinical variables for the study period (Multimedia Appendix 1).

Statistical Analyses

All statistical analyses were conducted using R, version 3.6.3 for Mac OS [35]. The following statistical packages were used: (1) dplyr [36] and tidyr [37] for data preparation; (2) lubridate [38] and stringr [39] for handling time and date; and (3) ggplot2 [40], patchwork [41], and scales [42] for plotting.

Descriptive Overview

For each hospital type, in addition to the number of hospitals, the total numbers of patients, admissions, and discharges were recorded by gender in frequencies and percentages. Length of stay (LOS), in days, was calculated for all patients by subtracting each individual's admission date from their discharge date. The results were aggregated at the level of the hospital type and presented as mean (95% CI) and median (IQR), for each type.

Capacity Utilization Measure

For each hospital, each study day's capacity utilization was calculated as a percentage, using that hospital's highest recorded daily bed occupancy for that year as the denominator [5,9]. Each patient's admission and discharge dates were used to calculate the number of patients present each day in each hospital. The capacity utilization of a day in a given hospital included all patients admitted or discharged on that day and the number who were present the previous day and not discharged or deceased. As noted above, patients admitted before the study year (eg, in December of the previous year) were included in the study until their discharge. Patients not discharged during the study year were included until the end of the study year. Total capacity utilization was calculated for each hospital (n=102) for 365 days (with a total of 37,230 time points).

Furthermore, daily percentages of capacity utilization were summarized by hospital type, along with means (95% CIs), SDs, and minimum-maximum values. To visualize daily variations in capacity utilization, smoothed lines were plotted, with the 95% CIs around the mean, for each hospital type. For each hospital type, we also plotted graphs to show variation by day of the week. Finally, for each hospital type, weekday vs weekend variations were plotted for each week of the year.

There was no dot plotted for weekdays of the first week, as the first day of this study year was a weekend (Sunday).

Patient Turnover Measure

As calculated for capacity utilization, daily patient turnover was calculated for each hospital and aggregated by hospital type. The patient turnover rate was calculated as absolute counts of admissions, discharges, and deceased patients for a particular day divided by the total number of patients for that day per hospital [21]. As opposed to using the inverse of the average LOS method, this approach has the advantage of adequately representing the volume of activity either for entire days or short hospital stays as contributors to increasing patient throughput [19,21,43]. The percentage of patient turnover per day was calculated and further summarized by hospital type as mean (95% CI), SD, and minimum-maximum figures. To display daily variations in patient turnover, similar displays were plotted for them as for capacity utilization.

Patient Clinical Complexity Level

Our patient clinical complexity level data covered only patients discharged during the study year (as the International Classification of Diseases [ICD]-10 codes were not available for patients until they were discharged), and Swiss DRG version 6 was applied for that study year [44]. The patient clinical complexity level calculation was based on a complex algorithm, providing clinical complexity and comorbidity level values (ranging from 0-4) for all possible primary and secondary diagnoses per patient case [26]. Developed as part of the complexity and comorbidity level Refinement Project in Australia [26,27,45], this algorithm was applied to determine each patient's final patient clinical complexity level. To facilitate this process, we used the grouping system provided by Swiss DRG AG [44].

We began by organizing our data input into a readable format via the grouping system. We chose the "SwissDRG Batchgrouper Format 2017" short input format, which provides an anonymous case identifier, plus the patient's age, sex, admission and discharge date, LOS, primary and secondary diagnoses, and treatment procedure codes. "DRG Output format for SwissDRG" results were then obtained, including patient clinical complexity level values for each case. Individual patient clinical complexity level values were further transformed to average daily patient clinical complexity level values per hospital, using each patient's admission and discharge dates (ie, each case's patient clinical complexity level value is applied to each day for that hospital until discharge). Daily patient clinical complexity level values were further summarized by hospital type as means (95% CIs), SDs, and minimum-maximum figures. As for the other 2 measures, similar displays were plotted to display variation in daily patient clinical complexity levels.

Results

Descriptive Overview

During the study year, 1,214,875 inpatients stayed in the 102 Swiss general hospitals, of which 16,379 cases (1.35%) continued to the following year. Of the 1,214,875 inpatients, the 5 university hospitals covered 19.50% (n=236,649), the 39 tertiary care hospitals covered 59.70% (n=725,317), and the 58 basic hospitals covered 20.81% (n=252,909) of the patient population. Overall, of the 1,214,875 inpatients, there were approximately 6.87% (n=91,078) more female than male patients. The average patient LOS across all general hospitals was 6.43 (95% CI 6.40-6.46) days; the median LOS was 3.7 (IQR 2.0-7.0) days. The general characteristics of the study population by hospital type are presented in Multimedia Appendix 1.

Variation of Daily Capacity Utilization

Average daily capacity utilization ranged from 527-2340 patients in university hospitals, 87-1099 patients in tertiary care hospitals, 16-304 patients in large basic hospitals, 7-179 patients in medium basic hospitals, and 1-93 patients in small basic hospitals. Notably, 3 small basic hospitals had average daily capacity utilization numbers below 10 patients.

Across the study period, the average daily capacity utilization was highest in university hospitals and the lowest in small basic hospitals (Table 1). However, the range (ie, the variation between the lowest and the highest daily capacity utilizations for a study year) was almost 98% (eg, 1.7-100.0) for small basic hospitals, 92% for medium basic hospitals, 87% for large basic hospitals, 73% for tertiary care hospitals, and 44% for university hospitals.

Table 1. Daily capacity utilization, patient turnover, and patient clinical complexity level per hospital, by hospital type, from the 1-year patient population.

Hospital type (median[a])	Capacity utilization (%)			Patient turnover[b] (%)			Patient clinical complexity level (0-4[c])		
	Mean (SD)	95% CI	Min-max[d]	Mean (SD)	95% CI	Min-max	Mean (SD)	95% CI	Min-max
University hospitals (988)	87.7 (7.7)	87.3-88.0	55.8-100.0	22.5 (7.6)	22.1-22.8	5.7-38.7	2.06 (0.2)	2.05-2.07	0.81-2.57
Tertiary care hospitals (298)	78.7 (10.2)	78.5-78.9	27.3-100.0	28.8 (7.5)	28.6-28.9	2.7-54.6	1.78 (0.3)	1.78-1.79	0.42-2.75
Large basic hospitals (120)	71.3 (13.4)	70.9-71.6	13.1-100.0	32.6 (9.2)	32.4-32.9	0.0-75.4	1.46 (0.4)	1.45-1.47	0.09-2.50
Medium basic hospitals (71)	65.3 (15.2)	65.0-65.6	5.9-100.0	34.5 (10.7)	34.3-34.7	0.0-109.1	1.26 (0.6)	1.25-1.27	0.00-2.93
Small basic hospitals (19)	57.8 (22.2)	57.3-58.4	1.7-100.0	24.3 (22.0)	23.8-24.9	0.0-200.0	1.65 (0.8)	1.63-1.67	0.00-4.00

[a] Median number of beds used per day per hospital, by hospital type.

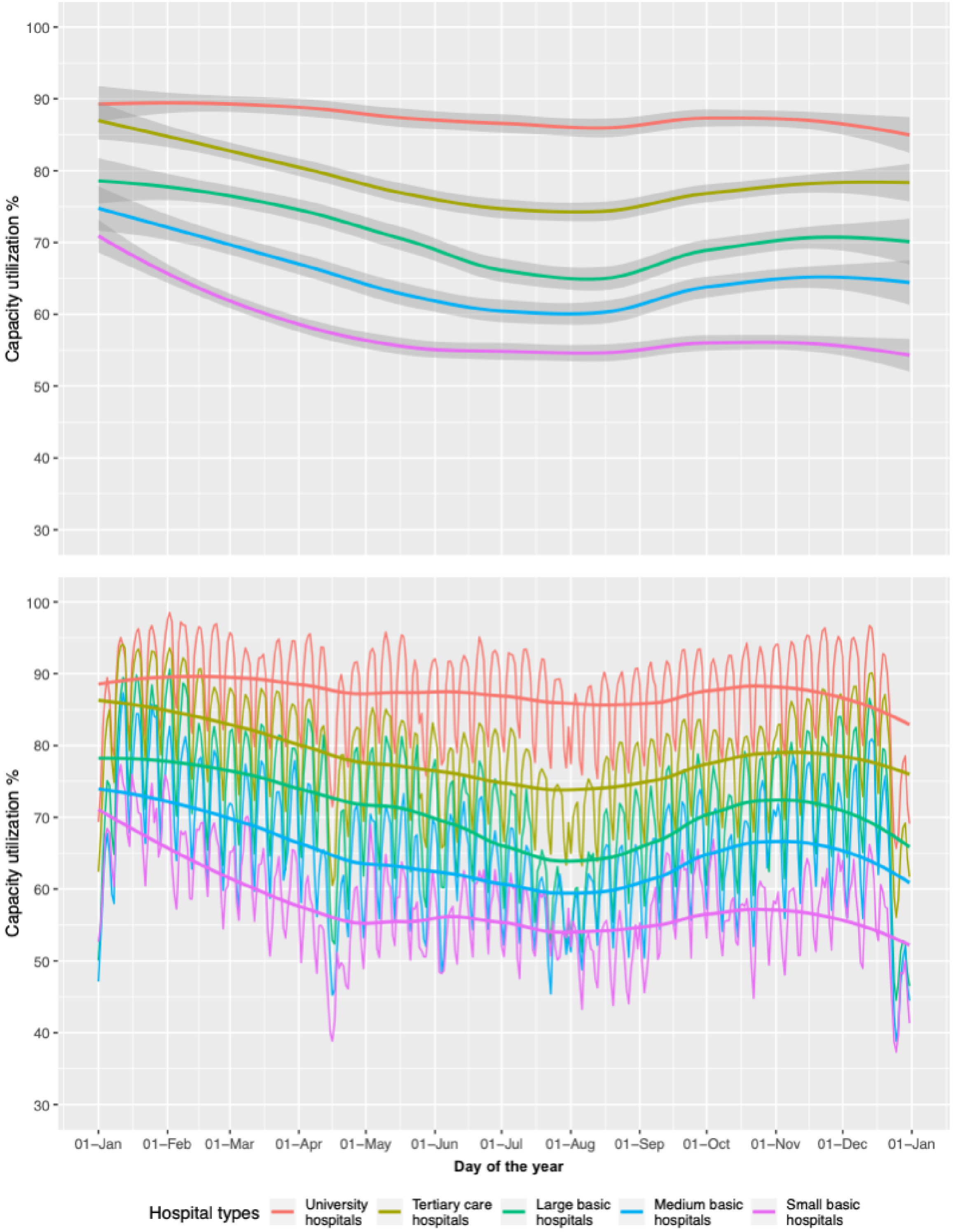
[b] Patient turnover is the percentage of total patients admitted and discharged in a day.

[c] 0=No clinical complexity, 1=mild clinical complexity, 2=severe clinical complexity, and 4=very severe clinical complexity.

[d] Min-max: Minimum-maximum value on hospital level within hospital type.

As indicated by the smooth curves and line charts by hospital categories, university hospitals' daily capacity utilizations were high throughout the year. Among all hospital types, capacity utilization was lower throughout the summer months (June-August; Figure 2).

Figure 2. Daily capacity utilization of Swiss general hospital types for 1 year (smooth curve with mean between CIs and line chart).



There was a gradual increase in capacity utilization through the early days of the week (Mondays-Wednesdays; Figure 3), followed by a gradual decrease from Fridays to Sundays, across all hospital types. There was roughly a 10% difference in capacity utilization during weekdays than on the weekend, in all hospital types (Table 2).

Figure 3. Capacity utilization of Swiss general hospital types with mean between CI by day of the week.

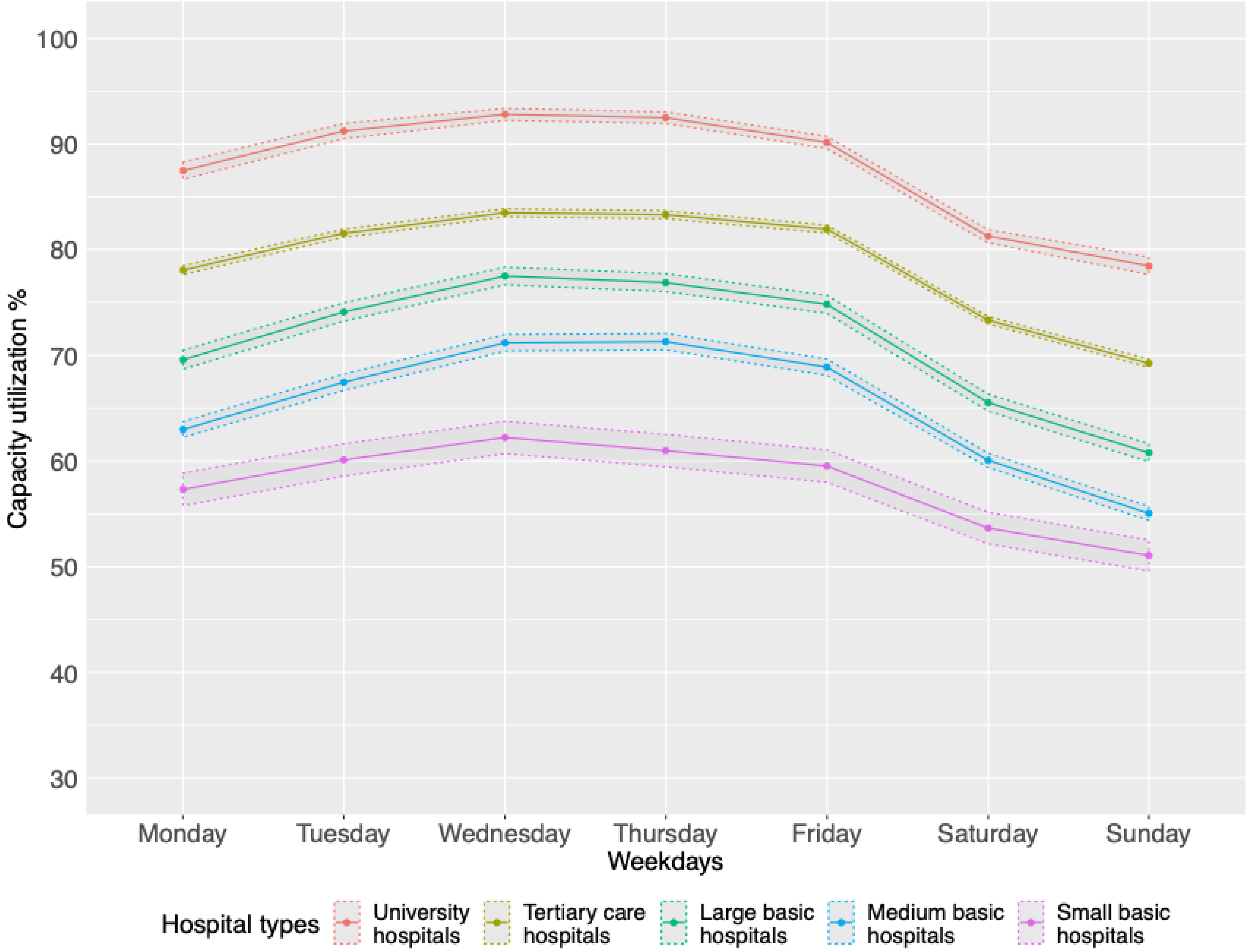


Table 2. Daily capacity utilization, patient turnover, and patient clinical complexity level by weekday vs weekend from the 1-year patient population.

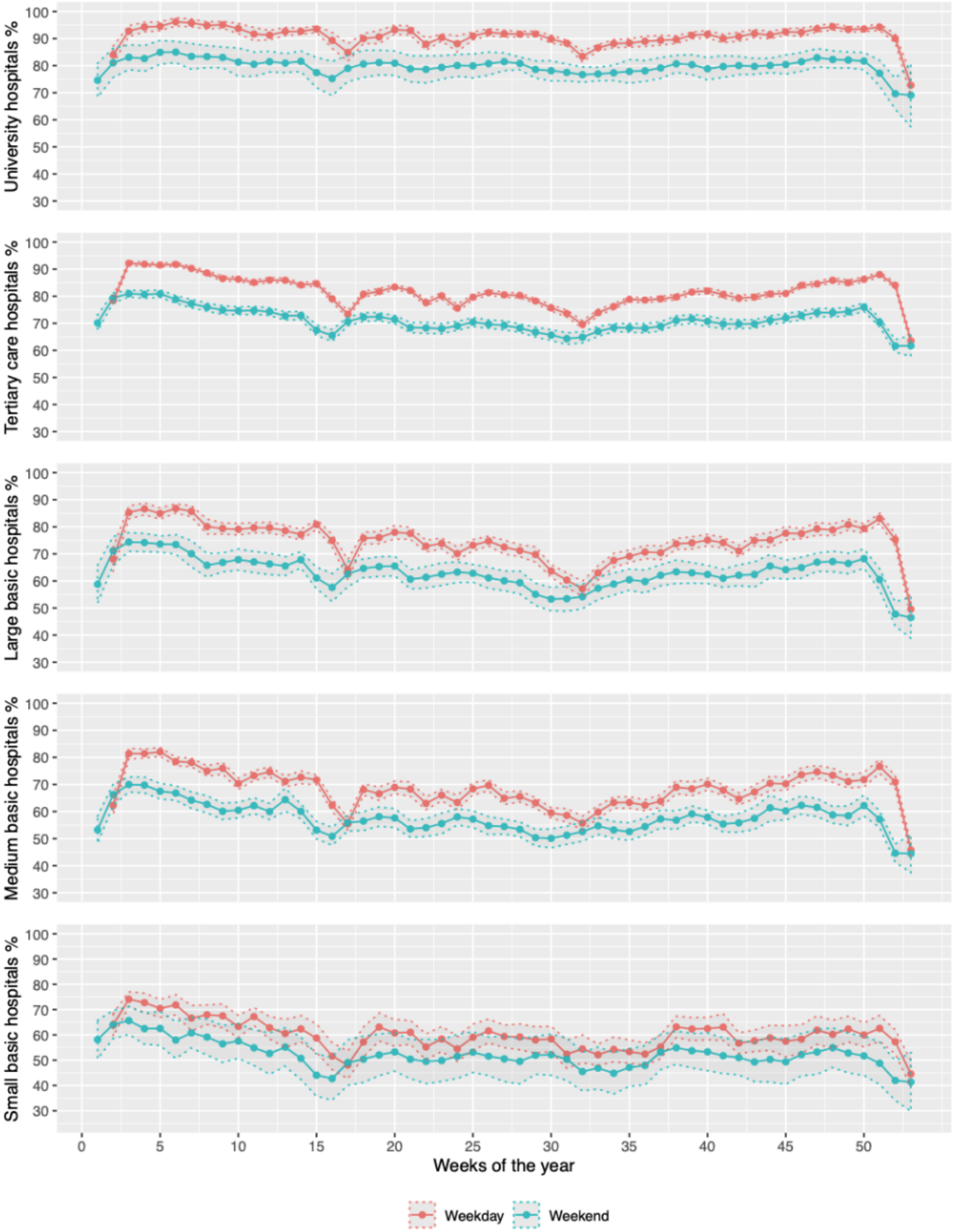
Day by hospital type	Capacity utilization (%)			Patient turnover (%)			Patient clinical complexity level (0-4[a])		
	Mean (SD)	95% CI	Min-max[b]	Mean (SD)	95% CI	Min-max	Mean (SD)	95% CI	Min-max
University hospitals									
Weekday	90.8 (3.8)	89.8-91.9	72.8-96.2	25.5 (1.4)	25.1-25.9	16.6-26.6	2.04 (0.21)	2.03-2.05	1.22-2.57
Weekend	79.8 (3.0)	78.9-80.6	69.1-85.0	14.8 (0.8)	14.6-15.1	10.5-17.6	2.1 (0.27)	2.08-2.13	0.81-2.55
Tertiary care hospitals									
Weekday	81.7 (5.6)	80.1-83.3	63.5-92.1	31.9 (1.6)	31.4-32.3	22.2-33.2	1.76 (0.28)	1.75-1.77	0.84-2.69
Weekend	71.2 (4.5)	69.9-72.4	61.6-80.9	21.1 (1.1)	20.8-21.4	14.4-24.3	1.83 (0.30)	1.82-1.84	0.42-2.75
Large basic hospitals									
Weekday	74.6 (7.4)	72.5-76.6	49.6-86.8	35.9 (1.9)	35.4-36.4	25.2-37.9	1.44 (0.38)	1.42-1.45	0.14-2.25
Weekend	63.0 (5.9)	61.4-64.6	46.5-74.4	24.4 (1.6)	23.9-24.8	16.7-29.5	1.52 (0.38)	1.5-1.54	0.09-2.5
Medium basic hospitals									
Weekday	68.4 (7.0)	66.4-70.3	45.8-82.0	37.1 (1.3)	36.8-37.5	31.2-39.3	1.24 (0.54)	1.22-1.25	0.05-2.79
Weekend	57.6 (5.5)	56.1-59.1	44.5-70.0	27.8 (1.7)	27.4-28.3	24.0-34.6	1.31 (0.57)	1.29-1.33	0.00-2.93
Small basic hospitals									
Weekday	60.0 (5.9)	58.3-61.6	44.6-74.1	27.0 (2.4)	26.3-27.6	22.2-31.7	1.63 (0.80)	1.6-1.65	0.00-4.00
Weekend	52.2 (5.3)	50.8-53.7	41.4-65.7	18.0 (4.0)	16.9-19.1	11.3-28.8	1.69 (0.81)	1.65-1.73	0.00-4.00

[a] 0=No clinical complexity, 1=mild clinical complexity, 2=severe clinical complexity, and 4=very severe clinical complexity.

[b] Min-max: Minimum-maximum.

Comparing weekdays with weekends, variations in capacity utilization for each week are shown in Figure 4. With very few exceptions (eg, the 2nd, the 17th, and the final week of the year), the weekly capacity utilization for weekdays was higher than for weekends. These weekly graphs also show lower capacity utilization during the summer months (weeks 20-35).

Figure 4. Capacity utilization of Swiss general hospital types with mean between CI by weekday vs weekend for 1 year.



Variation of Daily Patient Turnover

Throughout the year, daily patient turnover ranged from 72-468 patients for university hospitals, 5-318 patients for tertiary care hospitals, 0-91 patients for large basic hospitals, 0-78 patients for medium basic hospitals, and 0-33 patients for small basic hospitals. The minimum value of 0 indicates that some hospitals saw neither admissions nor discharges on some days.

During the study year, the mean daily patient turnover percentage was highest in medium-sized basic hospitals and lowest in university hospitals (Table 1). The difference in daily patient turnover range was highest for small basic hospitals, decreasing with each increase in hospital size class. And, as illustrated in the smooth and line chart, daily turnover varied the least in university hospitals and the most in medium basic hospitals (Multimedia Appendix 2).

Exploring the frequency of change in patient movement by day of the week (Multimedia Appendix 2), we found that turnover was highest on Mondays and lowest during weekends. Differences in patient turnover during weekdays and weekends across the 5 hospital types are shown in Table 2. Across the 5 hospital types, daily patient turnover was almost 10% higher on weekdays than on weekends.

The variation of patient turnover for each day of the week and for weekdays vs weekends is shown in Multimedia Appendix 2. In small basic hospitals, with few exceptions (eg, the 31st and 52nd weeks), the mean daily patient turnover was higher for weekdays than for weekends, for all hospital types. Overall patient turnover was lower during the holiday seasons (ie, weeks 17-18 and the final 2 weeks of the year).

Variation of Daily Patient Clinical Complexity Level

Overall, of the total discharged patients (1,198,496), 10.97% (n=131,442) of patients had severe clinical complexity (patient clinical complexity level 4), while roughly 60.97% (n=730,893) had no clinical complexity (patient clinical complexity level 0; Table 3).

Table 3. Patient clinical complexity level by hospital type (N=1,198,496).

Hospital type (N)	Patient clinical complexity level (0-4), n (%)				
	0: No clinical complexity, 730,893, (60.98)	1: Mild clinical complexity, 17,579 (1.47)	2: Moderate clinical complexity, 136,453 (11.39)	3: Severe clinical complexity, 182,129 (15.20)	4: Very severe clinical complexity, 131,442 (10.97)
University hospitals (n=232,127)	122,567 (52.8)	3281 (1.4)	28,812 (12.4)	41,159 (17.7)	36,308 (15.6)
Tertiary care hospitals (n=715,809)	433,514 (60.6)	11,003 (1.5)	83,374 (11.6)	110,756 (15.5)	77,162 (10.8)
Large basic hospitals (n=109,572)	73,800 (67.4)	1397 (1.3)	10,801 (9.9)	14,590 (13.3)	8984 (8.2)
Medium basic hospitals (n=123,530)	90,284 (73.1)	1679 (1.4)	11,237 (9.1)	12,865 (10.4)	7465 (6.0)
Small basic hospitals (n=17,458)	10,728 (61.5)	219 (1.3)	2229 (12.8)	2759 (15.8)	1523 (8.7)

Throughout the year, mean daily patient clinical complexity level varied across the 5 hospital types. It was highest in university hospitals (2.06, 95% CI 2.05-2.07) and lowest in medium basic hospitals (1.26, 95% CI 1.25-1.27; Table 1). This is depicted in the smooth and line chart for the 5 general hospital types (Multimedia Appendix 3). Mean patient clinical complexity level gradually decreased from Monday until midweek but remained highest during the weekend—the opposite of the usual pattern of capacity utilization and patient turnover explored by the days of the week (Multimedia Appendix 3).

Weekday and weekend differences in patient clinical complexity level for the 5 general hospital types are shown in Table 2. Across all hospital types, patient clinical complexity level was almost 0.07 points higher during the weekend than on weekdays. During weekdays, university hospitals' average daily patient clinical complexity level was 2.04 (95% CI 2.03-2.05); during the weekend, it was 2.1 (95% CI 2.08-2.13). Weekday vs weekend patient clinical complexity level variation over 1 year is shown in Multimedia Appendix 3. Except for a small number of weeks (eg, the 7th and the 31st weeks in small basic hospitals), across all hospital types, the patient clinical complexity level for weekends was higher than for weekdays. Moreover, except for small basic hospitals, patient clinical complexity level dropped in December. This was partly because patient clinical complexity levels could only be calculated for patients discharged during the study year (ie, ICD diagnostic codes were unavailable for patients not discharged during the year, and anonymity considerations made it impossible to track patients across years). However, mean values both for patient clinical complexity level and for LOS were also lower for patients discharged in November and December, with higher patient clinical complexity level values assigned to patients who remained in the hospital across the years' end (Multimedia Appendix 4).

Discussion

Principal Findings

We examined 1 year of routine patient data from all 102 general hospitals in Switzerland. Average daily capacity utilization varied widely, from 57.8% in small basic hospitals to 87.7% in university hospitals. However, patient turnover was highest, at 34.5%, in medium basic hospitals and lowest, at 22.5%, in university hospitals. Moreover, average daily patient clinical complexity level was highest in university hospitals, at 2.06, and lowest in medium basic hospitals, at 1.26. Surprisingly, in small basic hospitals, patient turnover was lower than in tertiary hospitals or either of the 2 other basic hospital types, both of which also had higher mean patient clinical complexity levels throughout the year. Another interesting finding was that the average daily patient clinical complexity level was highest on weekends. Additionally, all hospital types showed distinct weekday, weekend, and seasonal effects regarding capacity utilization, patient turnover, and patient clinical complexity level.

Concerns have been raised that capacity utilization alone does not explain hospitals' total care demand [6,18]. Therefore, we viewed this alongside daily volumes of admitted and discharged patients and complexity [18]. This study explored all 3 measures, showing that capacity utilizations and patient clinical complexity levels were highest but patient turnovers were lowest in university hospitals. Even with a large proportion of inpatients in tertiary care institutions, university hospitals generally operate at close to full capacity and with the most complex patient cases. Thus, more care resources need to be allocated to university hospitals [3]. On the other hand, in small basic hospitals, where capacity utilization and patient turnover were relatively low, patient clinical complexity level was above those of the other basic hospitals. This indicates that complex cases are still treated in small basic hospitals, possibly, due to geographic proximity, which may also relate to older patients' preference for them: across all hospital types, these hospitals have the highest mean patient age. In light of these small basic hospitals' continued relevance (as they still treat complex cases), they may warrant greater resource allocation.

We measured daily demand for Swiss general hospital care longitudinally for 1 year. As it has also been observed in other studies regarding days of the week, Saturdays and Sundays had the lowest capacity utilization and patient turnover [3,46]. Moreover, we observed that patient clinical complexity level was highest during weekends, possibly, because more complex patients remain in the hospital through the weekend. Comparing weekly demand throughout the year, a clear distinction between weekends and weekdays was shown, with the highest variability occurring in small basic hospitals, possibly indicating suboptimal patient flow. Concurrently, seasonal variations were also seen. Capacity utilization was mainly highest in the winter and relatively low in the summer months, whereas patient turnover was constant throughout the year, dropping off toward the end of the year. However, patient clinical complexity level remained quite constant, with a slight drop during the summer months and a marked reduction during December. These changes tended to correspond with holidays, possibly, including higher patient discharge rates and fewer admissions before the holidays and at the end of the year. The capture of daily patient complexity during the end of the year was also reduced because ICD-10 codes were unavailable for patients who were admitted but not discharged during the study year. Furthermore, patients with higher patient clinical complexity levels were more likely to have longer LOSs, particularly, across the Christmas and end-of-year period.

Potential Implications

Based on capacity utilization, patient turnover, and patient clinical complexity level, the variability of daily care demand in general hospitals directly impacts resource use. From the perspective of a single hospital, the extent of that impact depends on the degree of variation in care need, as well as on the hospital administrators' ability to adapt or otherwise respond to changes either in resource supply or demand. Our analysis on the demand dynamics of the Swiss health care system indicates that monitoring of care demand is useful to create surge capacity during disasters or the COVID-19 pandemic [20,47], by offering alternative solutions such as smoothing workloads and coordinating early discharges. It also has the potential to help health system planners and hospital managers to tailor their staffing and other resources to match care demand and the early planning of admissions (eg, surgeries or follow-up treatment) [48], to control patient flow for smoother service use.

What this analysis cannot describe is the human resources and other resources needed to meet care demands (ie, balancing care demands or any other supply-demand chain). To do so would require a full exploration of the relevant human resources (eg, physician and nurse staffing), in the light of each hospital's care demand. Furthermore, application of time-driven activity-based costing methods could provide a framework to identify process improvements for health care delivery [49,50]. However, we were not able to consider time-driven activity-based costing, as we do not have sufficient data regarding resource consumption (eg, personnel, equipment, and supplies) during the patients' journey along the clinical pathway [49].

Some studies have linked higher capacity utilization and patient turnover with adverse patient outcomes [5,21]. In addition to these results, acknowledging the effect of clinical complexity alongside capacity utilization and patient turnover might bring us closer to understanding the factors that stress hospital systems and the effects that a stressed system have on patient outcomes such as in-hospital mortality. Describing daily care demand to identify meaningful variation will require further studies (eg, examining patients' time-varying exposure to hospitals or units under pressure and the impact on the quality-of-care indicators and patient outcomes in causal models). Particularly, extending previous research on capacity utilization and in-hospital mortality [5,51] and using daily capacity utilization as time-varying exposure (ie, systemic stress factor), it would be of interest to explore the daily patient turnover and patient clinical complexity level as time-varying confounders. In a practical sense, this might also allow monitoring of pressure zones (eg, to manage care demand, where possible) in hospitals, which could reduce avoidable adverse events or death [5].

Strengths and Limitations

To our knowledge, this was the first study to explore hospital care demand dynamics via daily measurements of capacity utilization, patient turnover, and patient clinical complexity level on a national health system level. Furthermore, applying the standard methodology, programming, and software for large datasets allowed a longitudinal perspective by computing and visualizing demand dynamics per day of the year, day of the week, and weekdays vs weekends.

This study also had notable limitations. While we explored demand dynamics in detail, we could not do so with supply dynamics (eg, staffing, resources, etc)—an entire category of critical information in the demand-supply equation. Due to the large sample and FSO data composition (ie, aggregated data), it was also not possible to explore demand dynamics at the unit level—the interface between patients' care demand and health professionals' provision of care. Also, as we used codes assigned in routine data, the patient diagnoses and other variables could be biased by factors such as the accuracy of physicians' and nurses' documentation, lack of availability of ICD-10 codes for patients who were not discharged, and intentional upcoding of diagnoses to more expensive Swiss DRG categories [52,53].

Conclusions

This study illustrates daily care demand based on capacity utilization, patient turnover, patient clinical complexity level, and the variability of these factors between the 5 classes of Swiss general hospitals. For all 5 types, our analyses indicated distinct differences in capacity utilization, patient turnover, and patient clinical complexity level between days of the week, weekdays vs weekends, and seasons. This longitudinal study is a step toward detecting possible variables to be considered for time-varying exposure (eg, capacity utilization) and confounders (eg, patient clinical complexity level) in developing a casual model of tipping points and their links with quality of care or patient outcomes. Essentially, the variability of care demand provides a new perspective for gauging when hospitals are under strain, and this could help avoid pressure zones with a combination of appropriate resource allocation and care-demand planning in general hospitals.

Acknowledgments

We would like to thank the Swiss Federal Statistics Office (FSO) and its data management team for their support in providing inpatient data for our analyses. We are grateful to all the Swiss hospitals for their regular data reporting and information forwarded to the FSO. We would like to thank Mr Chris Shultis for English editing of this paper. We are also thankful to the Swiss Nurses' Association (SBK) for their financial support of this study. The funders had no role in the design, conduct, analysis, or report of this study.

Authors' Contributions

NS and MS had full access to all the data in the study and take responsibility for the integrity of the data and accuracy of the analysis. NS, RS, OE, DA, and MS contributed to the conception and design of the study. NS analyzed and drafted the manuscript. All the authors contributed to data interpretation and critically revised and edited the manuscript. All authors approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Description of variables and general characteristics of the study population for 5 Swiss general hospital types.

[PDF File (Adobe PDF File), 87 KB - jmir_v23i8e27163_app1.pdf]

Multimedia Appendix 2

Percentage of patient turnover by days of the year, days of the week (Monday to Sunday), and weekdays vs weekends over 1 year for 5 Swiss general hospital types.

[PDF File (Adobe PDF File), 498 KB - jmir_v23i8e27163_app2.pdf]

Multimedia Appendix 3

Average patient clinical complexity levels by days of the year, days of the week (Monday to Sunday), and weekdays vs weekends over 1 year for 5 Swiss general hospital types.

[PDF File (Adobe PDF File), 405 KB - jmir_v23i8e27163_app3.pdf]

Multimedia Appendix 4

Distribution of patient clinical complexity levels for the last 10 weeks of the year with the length of stay and average inpatients on weekly basis (showing complex patient stays over Christmas and the end of the year).

[PDF File (Adobe PDF File), 227 KB - jmir_v23i8e27163_app4.pdf]

References

1. Relman A. Book Review Crossing the Quality Chasm: A New Health System for the 21st Century By the Committee on Quality of Health Care in America of the Institute of Medicine. 337 pp. Washington, D.C., National Academy Press, 2001. $44.95. 0-309-07280-8. N Engl J Med 2001 Aug 30;345(9):702-703. [doi: 10.1056/nejm200108303450917] [Medline: 25057539]
2. Weber T. Optimal control theory with applications in economics. London: The MIT Press; 2011.
3. Belciug S, Gorunescu F. Improving hospital bed occupancy and resource utilization through queuing modeling and evolutionary computation. J Biomed Inform 2015 Feb;53:261-269 [FREE Full text] [doi: 10.1016/j.jbi.2014.11.010] [Medline: 25433363]

4. Griffiths P, Saville C, Ball J, Jones J, Pattison N, Monks T, Safer Nursing Care Study Group. Nursing workload, nurse staffing methodologies and tools: A systematic scoping review and discussion. Int J Nurs Stud 2020 Mar;103:103487 [FREE Full text] [doi: 10.1016/j.ijnurstu.2019.103487] [Medline: 31884330]
5. Kuntz L, Mennicken R, Scholtes S. Stress on the Ward: Evidence of Safety Tipping Points in Hospitals. Management Science 2015 Apr;61(4):754-771. [doi: 10.1287/mnsc.2014.1917]
6. Freedman S. Capacity and Utilization in Health Care: The Effect of Empty Beds on Neonatal Intensive Care Admission. Am Econ J Econ Policy 2016 May 01;8(2):154-185 [FREE Full text] [doi: 10.1257/pol.20120393] [Medline: 27942353]
7. Sloan F. Hsieh C-R. Health economics. London: MIT Press; 2017.
8. Jones RP. Condition-specific growth in occupied beds in England following a sudden and unexpected increase in deaths. British Journal of Healthcare Management 2019 Jun 02;25(6):1-8. [doi: 10.12968/bjhc.2018.0026]
9. Friebel R, Fisher R, Deeny SR, Gardner T, Molloy A, Steventon A. The implications of high bed occupancy rates on readmission rates in England: A longitudinal study. Health Policy 2019 Aug;123(8):765-772 [FREE Full text] [doi: 10.1016/j.healthpol.2019.06.006] [Medline: 31262535]
10. Reader TW, Gillespie A. Patient neglect in healthcare institutions: a systematic review and conceptual model. BMC Health Serv Res 2013 Apr 30;13:156 [FREE Full text] [doi: 10.1186/1472-6963-13-156] [Medline: 23631468]
11. Ball JE, Murrells T, Rafferty AM, Morrow E, Griffiths P. 'Care left undone' during nursing shifts: associations with workload and perceived quality of care. BMJ Qual Saf 2014 Feb;23(2):116-125 [FREE Full text] [doi: 10.1136/bmjqs-2012-001767] [Medline: 23898215]
12. Wagner J, Gabler NB, Ratcliffe SJ, Brown SES, Strom BL, Halpern SD. Outcomes among patients discharged from busy intensive care units. Ann Intern Med 2013 Oct 01;159(7):447-455 [FREE Full text] [doi: 10.7326/0003-4819-159-7-201310010-00004] [Medline: 24081285]
13. Panagioti M, Khan K, Keers RN, Abuzour A, Phipps D, Kontopantelis E, et al. Prevalence, severity, and nature of preventable patient harm across medical care settings: systematic review and meta-analysis. BMJ 2019 Jul 17;366:l4185 [FREE Full text] [doi: 10.1136/bmj.l4185] [Medline: 31315828]
14. OECD. Health at a Glance 2017. OECD Publishing. 2017. URL: https://www.health.gov.il/publicationsfiles/healthataglance2017.pdf [accessed 2020-07-08]
15. Fisher ES, Wennberg JE, Stukel TA, Skinner JS, Sharp SM, Freeman JL, et al. Associations among hospital capacity, utilization, and mortality of US Medicare beneficiaries, controlling for sociodemographic factors. Health Serv Res 2000 Feb;34(6):1351-1362 [FREE Full text] [Medline: 10654835]
16. Bed occupancy rate (%), acute care hospitals only. Published. 2018. URL: https://gateway.euro.who.int/en/indicators/hfa_542-6210-bed-occupancy-rate-acute-care-hospitals-only/ [accessed 2020-05-08]
17. Federal Statistical Office. Health Pocket Statistics 2017. Federal Statistical Office (FSO). URL: https://www.bfs.admin.ch/bfs/en/home/statistics/catalogues-databases/publications.assetdetail.4342090.html [accessed 2020-10-02]
18. Bujold M, Pluye P, Légaré F, Haggerty J, Gore GC, Sherif RE, Participatory Review Team. Decisional needs assessment of patients with complex care needs in primary care: a participatory systematic mixed studies review protocol. BMJ Open 2017 Nov 12;7(11):e016400 [FREE Full text] [doi: 10.1136/bmjopen-2017-016400] [Medline: 29133314]
19. VanFosson CA, Yoder LH, Jones TL. Patient Turnover: A Concept Analysis. ANS Adv Nurs Sci 2017;40(3):298-310. [doi: 10.1097/ANS.0000000000000171] [Medline: 28266962]
20. Chopra V, Toner E, Waldhorn R, Washer L. How Should U.S. Hospitals Prepare for Coronavirus Disease 2019 (COVID-19)? Ann Intern Med 2020 May 05;172(9):621-622 [FREE Full text] [doi: 10.7326/M20-0907] [Medline: 32160273]
21. Park SH, Blegen MA, Spetz J, Chapman SA, De Groot H. Patient turnover and the relationship between nurse staffing and patient outcomes. Res Nurs Health 2012 Jun;35(3):277-288. [doi: 10.1002/nur.21474] [Medline: 22457013]
22. Mendez CM, Harrington DW, Christenson P, Spellberg B. Impact of hospital variables on case mix index as a marker of disease severity. Popul Health Manag 2014 Feb;17(1):28-34 [FREE Full text] [doi: 10.1089/pop.2013.0002] [Medline: 23965045]
23. Menendez ME, Neuhaus V, van Dijk CN, Ring D. The Elixhauser comorbidity method outperforms the Charlson index in predicting inpatient death after orthopaedic surgery. Clin Orthop Relat Res 2014 Sep;472(9):2878-2886 [FREE Full text] [doi: 10.1007/s11999-014-3686-7] [Medline: 24867450]
24. Sharma N, Schwendimann R, Endrich O, Ausserhofer D, Simon M. Comparing Charlson and Elixhauser comorbidity indices with different weightings to predict in-hospital mortality: an analysis of national inpatient data. BMC Health Serv Res 2021 Jan 06;21(1):13 [FREE Full text] [doi: 10.1186/s12913-020-05999-5] [Medline: 33407455]
25. Quentin W, Geissler A, Busse R. Measuring and comparing health system outputs: using patient classification systems for efficiency analyses. In: Health system efficiency: European Observatory on Health Systems and Policies; 2016.
26. SwissDRG. Swiss Diagnosis Related Groups Version 8.0. Bern, Switzerland: SwissDRG AG; 2018. URL: https://www.swissdrg.org/de/akutsomatik/swissdrg [accessed 2020-12-05]
27. Bauer M, Ostermann H. DRGs in Transfusion Medicine and Hemotherapy in Germany. Transfus Med Hemother 2012 Apr;39(2):60-66 [FREE Full text] [doi: 10.1159/000337337] [Medline: 22670123]
28. Salisbury C, Boerma W. Balancing demand and supply in out-of-hours care. Florida: CRC Press; 2018.

29. Fieldston E, Ragavan M, Jayaraman B, Metlay J, Pati S. Traditional measures of hospital utilization may not accurately reflect dynamic patient demand: findings from a children's hospital. Hosp Pediatr 2012 Jan;2(1):10-18. [doi: 10.1542/hpeds.2011-0016-2] [Medline: 24319808]
30. Macy ML, Stanley RM, Lozon MM, Sasson C, Gebremariam A, Davis MM. Trends in high-turnover stays among children hospitalized in the United States, 1993-2003. Pediatrics 2009 Mar;123(3):996-1002 [FREE Full text] [doi: 10.1542/peds.2008-1428] [Medline: 19255031]
31. Unruh LY, Fottler MD. Patient turnover and nursing staff adequacy. Health Serv Res 2006 Apr;41(2):599-612 [FREE Full text] [doi: 10.1111/j.1475-6773.2005.00496.x] [Medline: 16584467]
32. Norrish BR, Rundall TG. Hospital restructuring and the work of registered nurses. Milbank Q 2001;79(1):55-79; 2 p preceding VI [FREE Full text] [doi: 10.1111/1468-0009.00196] [Medline: 11286096]
33. DeLia D. Annual bed statistics give a misleading picture of hospital surge capacity. Ann Emerg Med 2006 Oct;48(4):384-8, 388.e1. [doi: 10.1016/j.annemergmed.2006.01.024] [Medline: 16997673]
34. De Pietro C, Camenzind P, Sturny I, Crivelli L, Edwards-Garavoglia S, Spranger A, et al. Switzerland: Health System Review. Health Syst Transit 2015;17(4):1-288, xix [FREE Full text] [Medline: 26766626]
35. Antoch J. Environment for statistical computing. Computer Science Review 2008 Aug;2(2):113-122. [doi: 10.1016/j.cosrev.2008.05.002]
36. Wickham H, Francois R. dplyr: A Grammar of Data Manipulation. 2020. URL: https://cran.r-project.org/web/packages/dplyr/index.html [accessed 2020-08-28]
37. Wickham H. tidyr: Tidy Messy Data. 2020. URL: https://cran.r-project.org/web/packages/tidyr/index.html [accessed 2020-08-29]
38. Mailund T. Working with Dates: lubridate. New York: Springer; 2019:195-203.
39. Wickham H. stringr: Simple, Consistent Wrappers for Common String Operations. stringr: Simple, Consistent Wrappers for Common String Operations. 2019. URL: https://cran.rstudio.com/web/packages/stringr/index.html [accessed 2020-02-10]
40. Wickham H. ggplot2legant graphics for data analysis. New York: Springer; 2016.
41. Pedersen T. patchwork: The Composer of Plots. 2020. URL: https://cran.r-project.org/web/packages/patchwork/index.html [accessed 2020-12-18]
42. Wickham H. scales: Scale Functions for Visualization. 2020. URL: https://cran.r-project.org/web/packages/scales/index.html [accessed 2020-05-25]
43. Park SH, Dunton N, Blegen MA. Comparison of Unit-Level Patient Turnover Measures in Acute Care Hospital Settings. Res Nurs Health 2016 Jun;39(3):197-203. [doi: 10.1002/nur.21715] [Medline: 26998744]
44. SwissDRG. SwissDRG System 6. 2017. URL: https://www.swissdrg.org/de/akutsomatik/archiv-swissdrg-system/swissdrg-system-602017 [accessed 2020-05-04]
45. Antioch K, Zhang X. Organ transplant AN-DRGs: modifying the exceptions hierarchy in casemix classification. Aust Health Rev 2000;23(1):137-152. [doi: 10.1071/ah000137] [Medline: 10947598]
46. Pauls LA, Johnson-Paben R, McGready J, Murphy JD, Pronovost PJ, Wu CL. The Weekend Effect in Hospitalized Patients: A Meta-Analysis. J Hosp Med 2017 Sep;12(9):760-766. [doi: 10.12788/jhm.2815] [Medline: 28914284]
47. Mills AF, Helm JE, Wang Y. Surge Capacity Deployment in Hospitals: Effectiveness of Response and Mitigation Strategies. M&SOM 2020 Mar 30:367-387. [doi: 10.1287/msom.2019.0838]
48. Shabanikiya H, Gorgi HA, Seyedin H, Jafari M. Assessment of Hospital Management and Surge Capacity in Disasters. Trauma Mon 2016 May;21(2):e30277 [FREE Full text] [doi: 10.5812/traumamon.30277] [Medline: 27626015]
49. Martin JA, Mayhew CR, Morris AJ, Bader AM, Tsai MH, Urman RD. Using Time-Driven Activity-Based Costing as a Key Component of the Value Platform: A Pilot Analysis of Colonoscopy, Aortic Valve Replacement and Carpal Tunnel Release Procedures. J Clin Med Res 2018 Apr;10(4):314-320 [FREE Full text] [doi: 10.14740/jocmr3350w] [Medline: 29511420]
50. Keel G, Savage C, Rafiq M, Mazzocato P. Time-driven activity-based costing in health care: A systematic review of the literature. Health Policy 2017 Jul;121(7):755-763 [FREE Full text] [doi: 10.1016/j.healthpol.2017.04.013] [Medline: 28535996]
51. Madsen F, Ladelund S, Linneberg A. High levels of bed occupancy associated with increased inpatient and thirty-day hospital mortality in Denmark. Health Aff (Millwood) 2014 Jul;33(7):1236-1244. [doi: 10.1377/hlthaff.2013.1303] [Medline: 25006151]
52. Quan H, Eastwood C, Cunningham CT, Liu M, Flemons W, De Coster C, IMECCHI investigators. Validity of AHRQ patient safety indicators derived from ICD-10 hospital discharge abstract data (chart review study). BMJ Open 2013 Oct 10;3(10):e003716 [FREE Full text] [doi: 10.1136/bmjopen-2013-003716] [Medline: 24114372]
53. Wasserfallen J, Zufferey J. Financial impact of introducing the Swiss-DRG reimbursement system on potentially avoidable readmissions at a university hospital. Swiss Med Wkly 2015;145:w14097 [FREE Full text] [doi: 10.4414/smw.2015.14097] [Medline: 25659119]

Abbreviations

DRG: diagnostic-related group
FSO: Federal Statistics Office
ICD: International Classification of Diseases
LOS: length of stay

Edited by R Kukafka; submitted 13.01.21; peer-reviewed by B Zanotto, T Ricketts; comments to author 22.05.21; revised version received 17.06.21; accepted 05.07.21; published 19.08.21.

Please cite as:
Sharma N, Schwendimann R, Endrich O, Ausserhofer D, Simon M
Variation of Daily Care Demand in Swiss General Hospitals: Longitudinal Study on Capacity Utilization, Patient Turnover and Clinical Complexity Levels
J Med Internet Res 2021;23(8):e27163
URL: https://www.jmir.org/2021/8/e27163
doi:10.2196/27163
PMID:34420926

©Narayan Sharma, René Schwendimann, Olga Endrich, Dietmar Ausserhofer, Michael Simon. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Alignment of Key Stakeholders' Priorities for Patient-Facing Tools in Digital Health: Mixed Methods Study

Courtney Rees Lyles[1,2,3], PhD; Julia Adler-Milstein[4], PhD; Crishyashi Thao[4], MPH; Sarah Lisker[1,2], BA; Sarah Nouri[5], MD, MPH; Urmimala Sarkar[1,2], MD, MPH

[1]Division of General Internal Medicine, Department of Medicine, University of California San Francisco, San Francisco, CA, United States
[2]Center for Vulnerable Populations, Department of Medicine, University of California San Francisco, San Francisco, CA, United States
[3]Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA, United States
[4]Center for Clinical Informatics and Improvement Research, Department of Medicine, University of California San Francisco, San Francisco, CA, United States
[5]Division of Palliative Medicine, Department of Medicine, University of California San Francisco, San Francisco, CA, United States

Corresponding Author:
Courtney Rees Lyles, PhD
Division of General Internal Medicine
Department of Medicine
University of California San Francisco
1001 Potrero Avenue
Box 1364
San Francisco, CA, 94143
United States
Phone: 1 628 206 6483
Email: courtney.lyles@ucsf.edu

Abstract

Background: There is widespread agreement on the promise of patient-facing digital health tools to transform health care. Yet, few tools are in widespread use or have documented clinical effectiveness.

Objective: The aim of this study was to gain insight into the gap between the potential of patient-facing digital health tools and real-world uptake.

Methods: We interviewed and surveyed experts (in total, n=24) across key digital health stakeholder groups—venture capitalists, digital health companies, payers, and health care system providers or leaders—guided by the Consolidated Framework for Implementation Research.

Results: Our findings revealed that external policy, regulatory demands, internal organizational workflow, and integration needs often take priority over patient needs and patient preferences for digital health tools, which lowers patient acceptance rates. We discovered alignment, across all 4 stakeholder groups, in the desire to engage both patients and frontline health care providers in broader dissemination and evaluation of digital health tools. However, major areas of misalignment between stakeholder groups have stymied the progress of digital health tool uptake—venture capitalists and companies focused on external policy and regulatory demands, while payers and providers focused on internal organizational workflow and integration needs.

Conclusions: Misalignment of the priorities of digital health companies and their funders with those of providers and payers requires direct attention to improve uptake of patient-facing digital health tools and platforms.

(J Med Internet Res 2021;23(8):e24890) doi:10.2196/24890

KEYWORDS

medical informatics; medical informatics apps; information technology; implementation science; mixed methods

Introduction

While the US private sector has invested billions in digital health companies [1], and there have been rapid technological advancements [2], the majority of patients do not use digital health tools [3]. Broad public opinion surveys find that while 79% of Americans have searched for health information online, less than one-quarter of respondents had ever used patient-facing digital health tools, such as mobile device tracking or wearables [4,5]. Furthermore, the lowest rates of patient uptake of digital health tools are among underserved populations (such as low-income individuals) and chronically ill seniors [6,7]—these are populations with the highest overall burden of disease who might have the most to benefit from digital tools [5,8,9]. Finally, the impact of patient-facing digital health tools on clinical outcomes is not strong. Most privately developed patient-facing digital health tools lack an evidence base, and those with findings reported in peer-reviewed literature have mixed evidence of success [10-12]. Even where evidence exists, it often does not mature or scale; for example, despite evidence supporting the efficacy of diabetes self-management apps (such as those to assist and support users in tracking blood sugar levels, diet, and other behaviors), there are few long-term effectiveness studies [13], and few platforms have been implemented at scale across health care systems or nationwide.

In the midst of low adoption rates and limited broad scale evidence of effectiveness, investment in digital health companies continues to expand [1]—with a record $5.4 billion invested in the first half of 2020 during the recent COVID-19 pandemic [14]—underscoring the strong interest in and anticipated potential of using digital health tools to drive transformation. There have been large recent venture capitalist investments [15] specific to patient-facing digital health investment, such as in patient chronic disease management. The current state of digital health thus reflects an apparent disconnect between large financial investments in and population-level clinical and health gains from such tools [16]. However, little research has been performed to characterize and understand the root causes of this disconnect [17].

A deeper implementation-based understanding—of the priorities of private sector stakeholders' (who are driving much of the investment in digital health) and those of health care leaders (who are implementing digital health)—is critical to drive the broad use and impact of patient-facing digital health tools. Specifically, a better understanding of how stakeholders' perceptions of digital health priorities, opportunities, and gaps converge or diverge could serve to align the interests of key stakeholders in fostering digital health approaches that work across diverse populations and improve public and population health [17]. Therefore, we sought to systematically investigate key stakeholder groups' perceptions of adoption and effectiveness of patient-facing digital health tools.

Methods

Design

We pursued a 2-phase primary data collection process in which we conducted open-ended interviews and then conducted a focused structured survey among 4 key stakeholder groups in the digital health ecosystem. We collected qualitative and quantitative data on perceptions of the patient-facing digital health ecosystem. We defined patient-facing digital health tools as privately developed apps (and devices linked to apps, such as wearables) rather than electronic health record tools such as patient portals, which have followed a different implementation pathway, or broad educational websites such as MedlinePlus, which serve as a resource rather than an intervention platform. We sampled individuals in 4 stakeholder groups to reflect the investment or development perspective (venture capitalists and digital health companies) and the perspective of those purchasing and deploying the tools (payers, which included health plans and self-insured employers, and health care providers or leaders). While patients are another key audience in this ecosystem, there is a large body of evidence focused on patient interests in and barriers to using digital health tools [6,18-20]. We, therefore, add to the existing evidence base by focusing on the remaining stakeholders in this study. The University of California San Francisco Institutional Review Board approved this study (18-25418).

Conceptual Framework

We used an implementation science theoretical framework to identify areas of alignment and misalignment across stakeholder groups. The Consolidated Framework for Implementation Research (CFIR) comprises a set of domains designed to guide understanding of which practices, programs, or tools work and why across different contexts [21,22]. CFIR domains include the outer setting (events happening outside implementation, such as regulation), the inner setting (specific characteristics of health care organizations driving patient-facing digital health rollout), intervention characteristics (functionality and usefulness of the digital health tools, implementation processes, and individuals (patient or provider skills, knowledge, and beliefs). Given the multifactorial nature of the digital health ecosystem, CFIR allowed us to summarize data within and across multiple interacting domains and processes that could factor into successful implementation of patient-facing digital health tools and collectively influence uptake, effectiveness, and sustained use of patient-facing digital health.

Sample

We used expert networks in combination with snowball sampling to identify stakeholders for this study [23]. First, we identified experts and leaders within each stakeholder group and met regularly to brainstorm and compose the original participant outreach list. Then, we asked for additional contacts or recommendations from each interviewee. We sought to recruit 5 participants from each of the 4 stakeholder groups and for at least 2 out of the 5 participants to have expertise with Medicaid/Medicare or population health to ensure that issues pertaining to digital health equity were addressed. For example, we interviewed both safety net health plans and providers as well as those treating or focusing on privately insured populations and a mix of companies and venture capitalists with experience in chronic disease or population health-focused products.

Key Informant Interviews

We invited interview participants by email, and then scheduled and conducted individual hour-long interviews (by phone or video conference). The interview process lasted from March to December 2019, with 4 authors leading interviews; an additional author was present for notetaking. Participants provided verbal consent. All interviews were recorded and professionally transcribed.

Our semistructured interview guide (Multimedia Appendix 1) covered current patient engagement with digital health tools, factors affecting the development and adoption of patient-facing digital health solutions, health care system's climate for digital health implementation, and broad health policy and regulatory environment for digital health products. While the interview guide was structured broadly around the CFIR topics, we used open-ended topic exploration rather than specific framework domains or descriptions to drive the conversations. Furthermore, we specifically focused a portion of every interview on broad patient characteristics, such as age, socioeconomic status, and language, that have an impact on digital health tool use, given that there is literature on disparities among key patient groups [3,7,8].

Qualitative Analysis

Three members of the team coded 19 transcripts using the qualitative analysis software (Dedoose, version 9.0.17; SocioCultural Research Consultants LLC). The team first read the transcripts to deductively formulate codes mapped to top-level CFIR domains. Each CFIR domain was then examined across the 4 stakeholder groups, with an approach informed by a descriptive qualitative approach [24], given that there was no existing literature that directly compared perspectives among these diverse groups. Then, 2 team members individually recoded 4 of the transcripts using the refined codebook and inductive coding as new ideas emerged—often as concepts within each CFIR domain area. We referred any disagreements about coding to the entire study team for discussion to ensure that we followed a group-based process for data interpretation. Our coding collapsed inner setting and implementation processes into a single domain, given that there was variation among the stakeholders in the on-the-ground implementation experience of digital health tools (eg, venture capitalists had experience with workflows but would not have had extensive experience with individual health care system processes to elaborate on these concepts in depth). We also expanded coding to separately identify themes within the characteristics of individuals domain to generate information about both patient and provider perspectives, given that the beliefs and capabilities of both end users were described robustly during interviews. After finalization of the codebook, we divided and independently coded the remaining 15 transcripts.

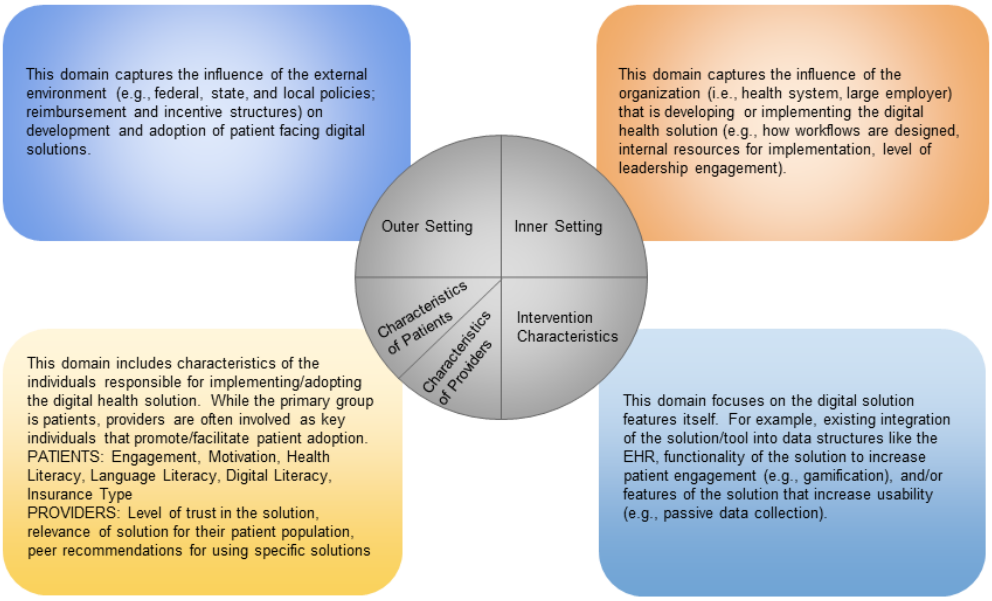
In the second phase of qualitative analysis, we used direct comparison of the codes and exemplar quotes across stakeholder groups in order to generate the overarching themes that cut across interview topics and CFIR domains. All team members were involved in this process, which also occurred iteratively as analysis unfolded.

Survey

We planned a second, structured phase of data collection to enable triangulation of the qualitative analysis. More specifically, to assess of the importance of each CFIR domain in the development and adoption of patient-facing digital health solutions, we developed a short survey. We sent the surveys and collected survey responses by email. Surveys were sent to all participants whom we previously interviewed, as well as to 1 additional digital health company and 4 additional payer organizations (to replace 4 interviewees who were unable to continue participating).

In the survey, we defined and explicitly named the CFIR domains that we used in the qualitative analysis: outer setting, inner setting, intervention characteristics, characteristics of the patient, and characteristics of the provider. Figure 1 displays the definitions shared with participants that explained CFIR domains and concrete examples of each domain from the qualitative analysis. We asked participants to allocate 100 points across CFIR domains on 2 survey items: (1) the extent to which the CFIR domain posed a challenge to the overall success and development of patient-facing digital health tools and (2) the extent to which the CFIR domain factored into the participant's daily professional decisions.

Figure 1. Domains that affect the development and adoption of patient-facing digital health solutions.



Outer Setting
Inner Setting
Characteristics of Patients
Characteristics of Providers
Intervention Characteristics

This domain captures the influence of the external environment (e.g., federal, state, and local policies; reimbursement and incentive structures) on development and adoption of patient facing digital solutions.

This domain captures the influence of the organization (i.e., health system, large employer) that is developing or implementing the digital health solution (e.g., how workflows are designed, internal resources for implementation, level of leadership engagement).

This domain includes characteristics of the individuals responsible for implementing/adopting the digital health solution. While the primary group is patients, providers are often involved as key individuals that promote/facilitate patient adoption.
PATIENTS: Engagement, Motivation, Health Literacy, Language Literacy, Digital Literacy, Insurance Type
PROVIDERS: Level of trust in the solution, relevance of solution for their patient population, peer recommendations for using specific solutions

This domain focuses on the digital solution features itself. For example, existing integration of the solution/tool into data structures like the EHR, functionality of the solution to increase patient engagement (e.g., gamification), and/or features of the solution that increase usability (e.g., passive data collection).

Survey Analysis

We calculated summary statistics (medians and interquartile ranges, given the small sample size and potential outlier responses) by CFIR domains. We compared stakeholder responses using radar charts and Mann–Whitney nonparametric tests.

Results

General

In total, we engaged 24 individuals across stakeholder groups, participants from 5 health-focused venture capitalists, 5 digital health companies, 8 payer organizations (including 3 focused on Medicaid and 4 commercial plans with a mix of private and Medicare/Medicaid products), and 6 providers (Table 1).

Table 1. Participant characteristics.

Stakeholder	Organization		
	Type	Region	Size (number of employees)
Venture capitalists			
Partner	Mature VC firm	West	51-100
Executive chairperson	Mature VC firm	West	11-50
Investor	Early-stage angel investor	West	≤10
Partner	Early-stage VC firm	West	11-50
Entrepreneur in residence	Early stage VC firm	West	≤10
Digital health companies			
Chief information officer	Series B company	West	11-50
Senior director of health plans	Series D company	West	251-500
Senior vice president[a]	Preseed company	West	≤10
Co-founder	Preseed company	Northeast	≤10
Senior vice president[b]	Seed company	West	11-50
Payer organizations			
Chief medical officer[a]	Medicaid health plan	West	251-500
Senior manager	Large employer organization	West	11-50
Chief medical officer[a]	Medicaid health plan	West	251-500
Senior vice president and chief digital officer[a]	Private insurance plan	Midwest	>10,000
Health informatics medical director[b]	Private insurance plan	West	11-50
Senior vice president[b]	Private insurance plan	West	5001-10,000
Chief executive officer[b]	Private insurance plan	Southeast	1001-5000
Chief medical officer[b]	Medicaid health plan	Multiple regions	5001-10,000
Health care providers or leaders			
Director of telehealth, specialty care provider	Safety net health care system	West	5001-10,000
Chief medical informatics officer, primary care provider	Safety net health care system	Midwest	5001-10,000
Clinic director, primary care provider	Safety net health care system	West	5001-10,000
Director of digital innovation, specialty care provider	Academic medical center	West	>10,000
Quality improvement lead, primary care provider	Academic medical center	West	>10,000
Director of biomedical informatics	Academic medical center	Northeast	1001-5000

[a] The individual participated in the interview only.

[b] The individual participated in the survey only.

Qualitative Interview Findings

Overview

Qualitative analysis findings of the in-depth interviews by CFIR domain, with a full set of exemplar quotes, are provided in Multimedia Appendix 2 (Tables S1-S4). There were some domains with alignment across stakeholder groups alongside those with significant misalignment of priorities and perceptions. Where misalignment was observed, it took the form of the perceptions and priorities of venture capitalists and digital health companies converging and those of payers and providers converging, with each pair expressing distinct viewpoints.

Outer Setting

All stakeholders described the need for changes to policy and the regulatory environment in order to drive widespread adoption of digital health tools. Current regulations were often cited as key barriers, for example,

> *The biggest challenge is really boring, but it's regulatory...the billing and coding systems.* [Digital health company]
>
> *[Our decision making] relates to HIPAA [Health Insurance Portability and Accountability Act] as a very, very conservative approach about privacy and access.* [Payer]

Stakeholders also agreed that current financial mechanisms and processes are mismatched with goals and needs of patients and frontline providers. One digital health company participant summarized the overall misaligned priorities in the external environment:

> *If you're a payer, you're really thinking about optimizing your cost. If you're a patient, you're really thinking about optimizing your health which may cost more, right? If you're a venture capitalist, you're really thinking about return on your investment. So, in that sense, everybody, all of the objectives are slightly at odds and [not] quite right, but they're not all aligned perfectly.*

Furthermore, venture capitalists and digital health companies emphasized the need for immediate financial return and the simultaneous desire to disrupt the system in a fundamental way. One venture capitalist participant stated,

> *If this [product] isn't something where you're trying to maximize margins or drive sales, drive revenue, then it might make sense for you to think about other forms of capital.*

Similarly, a participant from a digital health company stated,

> *There's a lot of pressure to grow very fast, and it doesn't take into account the mission of the organization, because of what the [financial] opportunity is.*

In contrast, while payers and providers mentioned the need for external incentives to drive patient-facing digital health uptake, they commented in a more balanced way on the need for digital products that improve incrementally as well as those that are disruptive. One provider summarized this, saying,

> *There are some companies that are aligned with trying to sell into what our business model is and then there are others who are trying to reinvent health care in terms of how health care is delivered, how health care is paid for.*

Inner Setting

Stakeholders agreed that frontline perspectives from patients and providers are valuable but not often obtained. In addition, most stakeholder groups pointed to the need to improve integration of patient-facing digital tools into provider workflows. For example, interviewees stated,

> *I think we all know that it is important to also make sure that you're educating and getting feedback from the boots on the ground that are actually going to be using it, but...it's really easy to just focus on getting the contract signed and getting the executive buy-in.* [Venture capitalist]
>
> *I think a lot of the digital health and IT [information technology] implementation...is more trickledown... At least in my experience, I don't think that there has been much of an outreach to get provider opinion about these things before going in.* [Health care provider]

However, within this domain, the providers particularly emphasized lack of provider workflow integration as key barrier:

> *The reality is no one wants to open a second program to try to do something when they're busy and they're trying to do things...Maybe even more so for providers, for doctors, nurse practitioners and people like that who have hard-to-change behavior already.*

Finally, payers and providers also offered concrete examples of bandwidth and staffing challenges to advancing patient adoption of digital tools. For example, interviewees stated,

> *We have bandwidth issues. We have things that we have to do because the state tells us, or some of our big providers...say it's a priority for them. It's being able to eke out enough space to work on something that might well be considered discretionary.* [Payer]
>
> *There may be negative bandwidth to do that kind of stuff...this small piece of integration to get the reports in our [EHR] system.* [Health care provider]

Intervention Characteristics

With respect to the characteristics of the digital health tools themselves, most interviewees mentioned the need for better functionality, particularly in terms of ease of use and seamless data sharing.

> *Number one, it has to be incredibly simple and intuitive.* [Digital health company]
>
> *[There is a lack of] a seamless end-to-end continuum [of data sharing/transmission] that can enable a person's longitudinal health over time.* [Payer]

However, the largest differences between stakeholders within this domain were related to evidence about the uptake and effectiveness of the digital health tools. Payers and providers particularly emphasized the lack of sufficient evidence to show that large groups of patients would use available digital tools, as well as a lack of evidence demonstrating clinical benefits:

> *The [self-insured] employers say, "Almost no one's using them [digital health tools]," and then the vendor [digital health company] is saying, "Well, yes, our engagement rates are 75%." So I think there's a big disconnect between what the vendors think is possible and what the reality is of the employees either finding or wanting to use these tools.* [Payer]
>
> *If we put this app there as a benefit, you got to feel reasonably confident it's going to have...a reasonable likelihood to benefit. What's the evidence for that?* [Payer]

Characteristics of Individuals

Finally, when considering both patient and provider characteristics, all groups mentioned their desire to focus meaningfully on patient engagement with digital health tools to make the largest impact in the digital health ecosystem. For example, many interviewees mentioned patient age (ie, older patients) and lower behavioral readiness as factors inhibiting use, suggesting that certain subgroups of patients need targeted approaches to achieve broad uptake. For example, interviewees stated,

> *These apps are still on the whole for techies, so it's hard when you start getting into populations like the whole senior population.* [Venture capitalist]
>
> *[If] the [patients] that don't want to engage at all, the solution obviously cannot reflect the need that they have, and that's the gap that we've seen.* [Payer]

Most interviewees also suggested that patient uptake depends on provider engagement and recommendation of digital health tools. Despite this, they offered few examples or concrete strategies for how to work with providers to promote patient engagement. For example, interviewees stated:

> *If the patient knows that someone or that their doc is following along, they're more likely to remain engaged.* [Digital health company]
>
> *I tend to avoid giving patient recommendations [for any app or digital platform], except for a few things I've specifically looked at where I trust the source.* [Health care provider]

Overall, venture capitalists and digital health companies described generic patient interest and ability to use digital health tools, while the providers in particular gave specific examples of patient factors that impede digital platform use, such as digital and health literacy and language accessibility. This contrast was exemplified in quotations:

> *I think making the individual who has the chronic condition the center of all of the efforts is key... [Consider] the barriers to people living in the most healthful way possible.* [Digital health company]
>
> *I think English speaking is a pretty big piece of that because most of the apps are not going to be available in necessarily multiple languages, certainly not beyond probably Chinese, Spanish—probably not even beyond Spanish.* [Health care provider]

Overarching Qualitative Themes

Synthesizing across CFIR domains revealed 3 themes—(1) Patient needs and preferences are secondary, (2) lack of shared definition of success blocks progress, and (3) each stakeholder group focuses on immediate but diverging priorities.

Patient Needs and Preferences Are Secondary

While all stakeholder groups believe it is important for the digital health sector to design for patients who are diverse in backgrounds, needs, and preferences, few put this into practice and consider how tools should be altered to accommodate varied patient characteristics. Venture capitalists and digital health companies focused more on outer setting (eg, reimbursement, regulation) and intervention characteristics (eg, app functionality) for mass market dissemination. While payers recognized barriers to patient use of digital health tools, they also lacked strategies to achieve meaningful patient engagement. Finally, while providers were best able to articulate the needs of individual patients (such as variation by age, socioeconomic status, and digital literacy), they had little bandwidth or ability to recommend digital tools to their patients or integrate digital tool use in practice.

Lack of Shared Definition of Success Blocks Progress

There is a lack of evaluation of digital health tools, with payers and providers stating that there are no clear patient-facing tools to recommend based on evidence. Underscoring this is a misalignment between stakeholder groups' definitions of digital health platform effectiveness or success. Differing definitions of success can lead to different stakeholders focusing on different outcomes, which divides focus and therefore impedes progress. This was evident when digital health companies focused on high-level utilization data, such as the total number of downloads or users of their platform, while providers and payers focused on behavioral or outcome measures, including who was offered versus who adopted the technology and examination of adoption overall, as well as, by key patient groups (eg, uptake of tools only among younger, healthier populations who are already engaged in the healthcare system, also known as the "worried well").

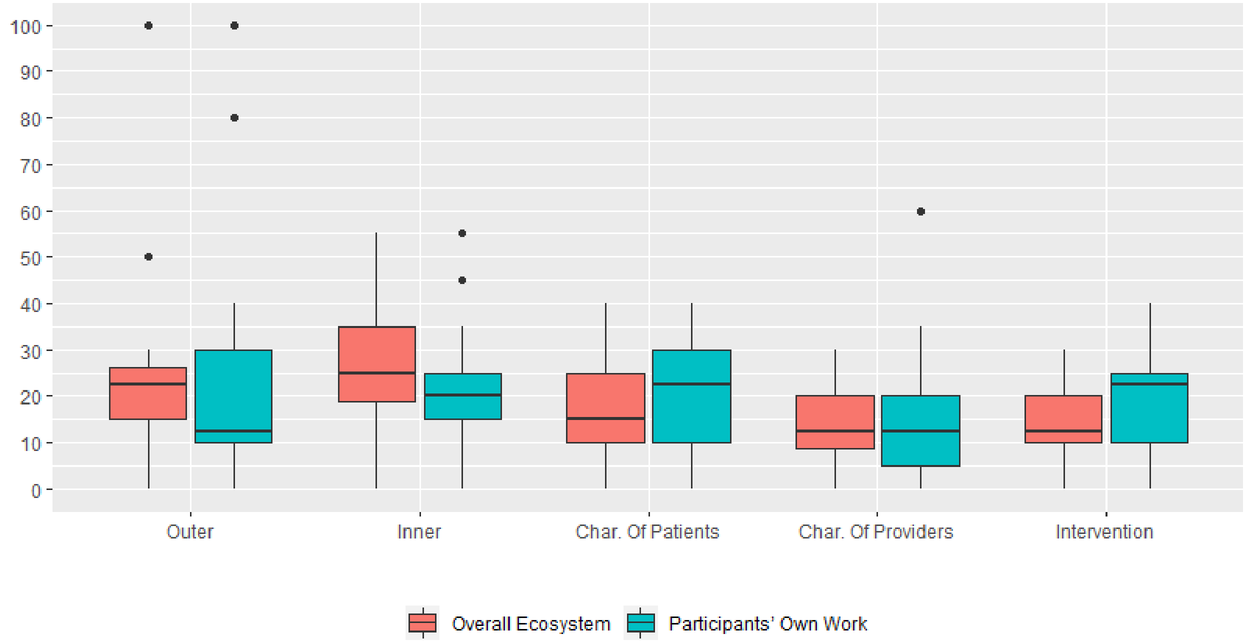
Each Stakeholder Group Focuses on Immediate but Diverging Priorities

There are large differences between stakeholder groups in how they view the most pressing needs within the patient-facing digital health space. For providers, it was clear that lack of integration into electronic health record systems and workflows are a huge challenge to adoption. This inner setting challenge, in turn, leads providers to be less willing to recommend tools to their patients and want to communicate about use of digital tools in the context of their existing care, missing an opportunity to drive patient engagement. Digital health companies and venture capitalists focused most on the outer setting challenges such as regulatory compliance including billing and coding systems, privacy, and incentives that prevent uptake of digital health tools.

Quantitative Survey Findings

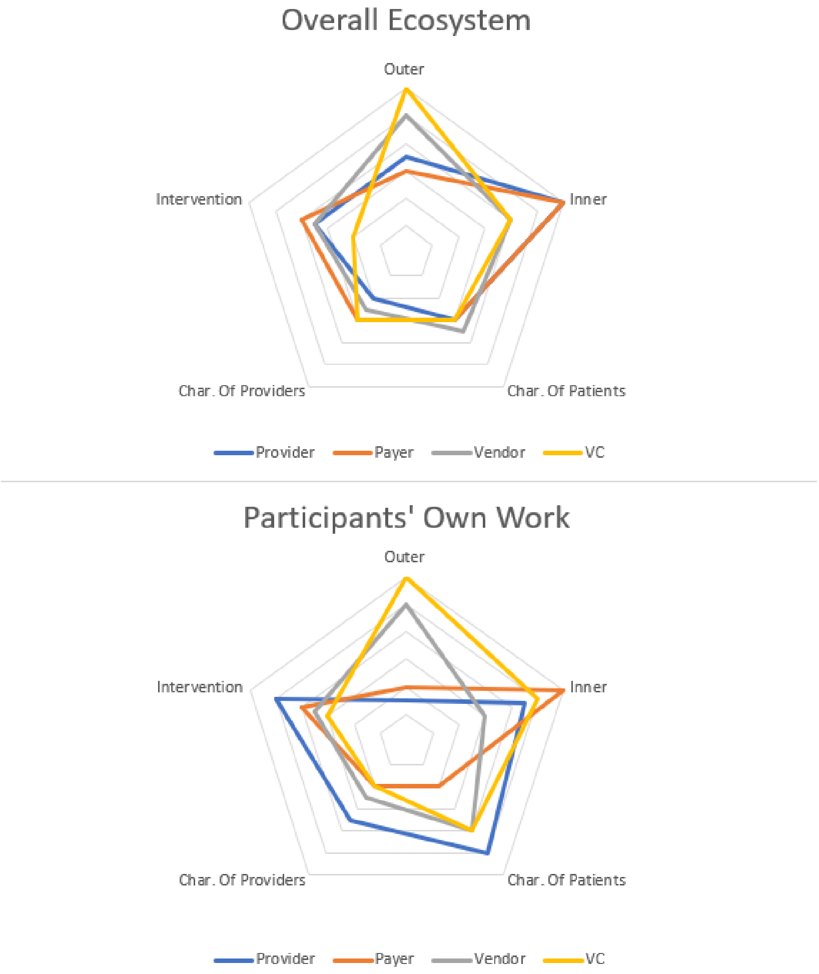
When asked to consider the CFIR domains in a structured, comparative way via survey, all CFIR domains were perceived as important drivers of the patient-facing digital health ecosystem. Point allocation among participants was highest for the outer setting and inner setting in the overall ecosystem. Scores differed somewhat when we asked for allocations in the participants' own work, with slightly higher median scores in the inner setting, characteristics of patients, and intervention domains (Figure 2).

Figure 2. Responses by Consolidated Framework for Implementation Research domain. Char: Characteristics.



We found evidence of stakeholder misalignment that was consistent with findings from the interviews. From radar charts (Figure 3), we observed that digital health companies and venture capitalists were aligned in their assessment, and emphasized outer setting as posing a challenge to the patient-facing digital health ecosystem and as the domain with the largest effect on their individual work. Providers and payers were aligned in their assessment of the importance of the inner setting in the overall patient-facing digital health ecosystem, yet providers were most likely to focus on patient characteristics in their own work. Venture capitalists had significantly lower scores for the intervention domain within the digital health ecosystem than payers (*P*=.02) and providers (*P*=.03). Providers had significantly lower scores for the outer setting domain within the individual work question than venture capitalists (*P*=.049) and digital health companies (*P*=.02).

Figure 3. Radar chart showing responses by stakeholder group. Char: Characteristics; VC: venture capitalist.



Discussion

Principal Findings

While private-sector digital health solutions that are implemented by health plans and health care delivery systems are widely touted as a key driver of health system transformation [25,26], our study revealed overlap in some conceptual domains along with clear gaps between these sectors. All stakeholder groups (venture capitalists, digital health companies, payers, and providers) identified frontline provider, staff, and patient engagement as drivers of widespread uptake of patient-facing digital tools. Yet our analyses of key informant interviews and survey results revealed misalignment between the focuses of these groups that are likely impeding the ability to achieve that uptake. Specifically, venture capitalists and digital health companies often focused on issues in the outer setting (eg, regulation). In contrast, provider and payer interviews largely focused on issues in the inner setting (eg, workflows). These overall emphasis areas were maintained in our follow-up surveys with stakeholders as they ranked the priorities of various implementation domains in their daily work. Taken together, as we synthesized findings into broad themes, we found that patient characteristics (eg, needs and preferences related to digital health tool use) were often addressed secondarily in the digital health ecosystem.

Our work sheds light on the specific nature of the disconnect within this ecosystem, on which other studies have suggested similar concepts. For example, recent expert synthesis to advance the digital health landscape has also called for stronger evidence of clinical improvements from digital health apps and more careful consideration of workflow integration to drive widespread implementation [27]. Similarly, consensus reports from stakeholder convenings, such as reports from influential organizations like the World Health Organization [17], also support the need for bringing diverse stakeholders together to set and carry out shared priorities. Furthermore, a large body of patient-facing research on barriers to digital health tool use highlights digital literacy barriers and the desire to use technology that augments their existing provider and staff relationships [6], which is consistent with our findings. Our study makes a unique contribution by collecting and analyzing primary data from a diverse set of stakeholders and focuses on multiple rather than a single stakeholder group [28-31]; therefore, our analyses add detail to the problems discussed in the field, informed by an implementation framework, that could advance our next steps in a multilevel fashion.

Policy and Practice Implications

Moving forward, the diverging priorities across stakeholder groups will make it difficult to implement, spread, and sustainably reimburse patient-facing digital health tools in real-world health care settings. Our work suggests a role for increased communication among these stakeholders, as each stakeholder group lacked detailed understanding of other groups' immediate priorities. If federal and state policy could remove barriers in the outer setting, then venture capitalists and companies may be better able to anticipate regulatory environments that protect patients yet move and adapt at the appropriate pace to support new innovation. Recent evidence suggests major players such as the Federal Drug Administration and the Centers for Medicare and Medicaid Services are fostering change in this space [32]. These efforts by key policy stakeholders have promoted the growth of digital tools, primarily by increasing patients' ability to access their health data in electronic formats that can be connected to smartphones and other platforms. In parallel, these entities have also pursued more oversight over tools—in particular those that act as medical devices to help ensure safety and efficacy. In turn, these efforts could promote consumer confidence and broader adoption. However, the outer setting issues identified related to reimbursement and regulation were broad—spanning from the need for broader value-based payment to privacy or security challenges to regulation of novel technologies (eg, artificial intelligence). While federal policymakers are tackling each of these areas, these efforts are not guided by a singular focus on advancing digital health and so the results are likely to be uneven.

Moreover, these outer setting changes will likely not succeed without simultaneous focus on inner setting barriers, such as integration into workflows and increased focus on provider and patient needs [33]. Even in light of the COVID-19 pandemic, during which many patient-facing digital tools are rapidly being tested to support patients remotely, a dearth of workflows to support their use—coupled with lack of evidence generation about health improvements and exacerbation of health disparities—could continue to suppress wide scale adoption in the long term [34,35]. In particular, business models for digital health companies will diverge from frontline health care system needs until these stakeholders can tie return on investment to meaningful and immediate priorities for patient care. Therefore, direct work across stakeholders to create a shared agenda for working more closely together will be critical to create alignment. All stakeholders must be willing to jointly decide (from pilots to large scale implementation) on (1) reimbursement/business models that incentivize diverse patient uptake of tools; (2) standard workflows and processes for piloting tools within real-world settings; (3) evaluation metrics that address uptake, engagement, and clinical effectiveness; and (4) policy and regulatory oversight that maximizes speed while maintaining quality and safety of digital approaches.

Limitations

There are several limitations of this study that are important to note. First, the sample size was modest and overrepresents stakeholders in the western United States and academic medical centers in urban and suburban areas (likely driven by both the sampling strategy and the geographic clustering of venture capitalist firms and digital health companies). Furthermore, we did not explicitly sample digital health researchers or patients in this study, given our decision to focus on privately developed digital health tools and stakeholder groups that were less represented in previously published literature. Future work with a broader sampling approach across all stakeholder groups is needed. However, our mixed methods approach likely increased the potential generalizability of our findings. In addition, we focused our analysis on top-level CFIR domains rather than on CFIR subdomains, given the large misalignment that was evident at the highest domain levels in our qualitative coding. Future work is needed to flesh out nuanced barriers and facilitators at the subdomain level.

Conclusions

Despite the presence of some overlapping perspectives across stakeholder group priorities, the gaps and misalignment between digital health companies and their funders, on one hand, and providers and payers, on the other, deserves direct attention in striving for digital transformation. Closer, longitudinal collaboration among stakeholders and team-based approaches may address this fundamental challenge—especially to ensure that digital solutions are matched to the needs of the diverse US population [36].

Acknowledgments

This work was supported by The Commonwealth Fund (grant 20181504). The views presented here are those of the authors and not necessarily those of The Commonwealth Fund, its directors, officers, or staff. CRL and US were also funded by University of California San Francisco S.O.L.V.E. Health Tech, with contract funding from InquisitHealth (CA-0146605), AppliedVR (CA-0143878), and SomnologyMD (CA-0153574). No funders had any part in data analyses or final results presented herein.

Authors' Contributions

CRL and JAM conceived of and designed the study. CRL, JAM, CT, and US conducted qualitative interviews, and all authors conducted qualitative and quantitative analyses. CRL wrote the manuscript with critical revisions from all authors. All authors approved the final submitted version and take responsibility for the integrity of this work.

Conflicts of Interest

CRL and US co-lead a university initiative, UCSF S.O.L.V.E. Health Tech, that holds contract funding from InquisitHealth, AppliedVR, and SomnologyMD; SL is supported by these funds as well. JAM is on the board of Project Connect. US serves as an expert advisor for nonprofit digital health organizations HealthTech 4 Medicaid and HopeLab, is a member of the American

Medical Association’s Equity and Innovation Advisory Group, has been a clinical advisor for Omada Health, and has been an advisory board member for Doximity. These groups played no role in the study design, data collection, analysis, or presentation of results.

Multimedia Appendix 1
Interview guide.
[DOCX File , 26 KB - jmir_v23i8e24890_app1.docx]

Multimedia Appendix 2
Example quotations (by CFIR domain).
[DOCX File , 42 KB - jmir_v23i8e24890_app2.docx]

References

1. Day S, Gambon E. In 2019, digital health celebrated six IPOs as venture investment edged off record highs. Rock Health. 2019. URL: https://rockhealth.com/reports/in-2019-digital-health-celebrated-six-ipos-as-venture-investment-edged-off-record-highs/ [accessed 2021-08-02]
2. Steinhubl SR, Muse ED, Topol EJ. The emerging field of mobile health. Sci Transl Med 2015 Apr 15;7(283):283rv3. [doi: 10.1126/scitranslmed.aaa3487] [Medline: 25877894]
3. Mahajan S, Lu Y, Spatz ES, Nasir K, Krumholz HM. Trends and predictors of use of digital health technology in the United States. Am J Med 2021 Jan;134(1):129-134. [doi: 10.1016/j.amjmed.2020.06.033] [Medline: 32717188]
4. Day S, Seninger C, Kothari M, Fan J, Pundi K, Perino A, et al. Digital health consumer adoption report. Rock Health. 2019. URL: https://rockhealth.com/reports/digital-health-consumer-adoption-report-2019/ [accessed 2021-08-02]
5. Zweig M, Shen J, Jug L. Health care consumers in a digital transition. Rock Health. 2017. URL: https://rockhealth.com/reports/healthcare-consumers-in-a-digital-transition/ [accessed 2021-08-02]
6. Lyles CR, Nelson EC, Frampton S, Dykes PC, Cemballi AG, Sarkar U. Using electronic health record portals to improve patient engagement: research priorities and best practices. Ann Intern Med 2020 Jun 02;172(11 Suppl):S123-S129 [FREE Full text] [doi: 10.7326/M19-0876] [Medline: 32479176]
7. Lin SC, Lyles CR, Sarkar U, Adler-Milstein J. Are patients electronically accessing their medical records? evidence from national hospital data. Health Aff (Millwood) 2019 Nov;38(11):1850-1857. [doi: 10.1377/hlthaff.2018.05437] [Medline: 31682494]
8. Nouri S, Adler-Milstein J, Thao C, Acharya P, Barr-Walker J, Sarkar U, et al. Patient characteristics associated with objective measures of digital health tool use in the United States: a literature review. J Am Med Inform Assoc 2020 May 01;27(5):834-841 [FREE Full text] [doi: 10.1093/jamia/ocaa024] [Medline: 32364238]
9. Vogels EA. Digital divide persists even as lower-income Americans make gains in tech adoption. Pew Research Center. 2021 Jun 22. URL: https://pewrsr.ch/2TRM7cP [accessed 2021-08-02]
10. Safavi K, Mathews SC, Bates DW, Dorsey ER, Cohen AB. Top-funded digital health companies and their impact on high-burden, high-cost conditions. Health Aff (Millwood) 2019 Jan;38(1):115-123. [doi: 10.1377/hlthaff.2018.05081] [Medline: 30615535]
11. Chauvin J, Lomazzi M. The digital technology revolution and its impact on the public's health. Eur J Public Health 2017 Dec 01;27(6):947. [doi: 10.1093/eurpub/ckx134] [Medline: 29088447]
12. Sarkar U, Lyles C. Devil in the details: understanding the effects of providing electronic health record access to patients and families. BMJ Qual Saf 2020 Dec;29(12):965-967. [doi: 10.1136/bmjqs-2020-011185] [Medline: 32699029]
13. Fleming GA, Petrie JR, Bergenstal RM, Holl RW, Peters AL, Heinemann L. Diabetes digital app technology: benefits, challenges, and recommendations. a consensus report by the European Association for the Study of Diabetes (EASD) and the American Diabetes Association (ADA) Diabetes Technology Working Group. Diabetes Care 2020 Jan;43(1):250-260. [doi: 10.2337/dci19-0062] [Medline: 31806649]
14. Kramer A, Shah A. 2020 midyear digital health market update: unprecedented funding in an unprecedented time. Rock Health. 2020. URL: https://rockhealth.com/reports/2020-midyear-digital-health-market-update-unprecedented-funding-in-an-unprecedented-time/ [accessed 2021-08-02]
15. Hathaway I, Rothwell J. A cure for health care inefficiency? the value and geography of venture capital in the digital health sector. The Brookings Institution. 2015 Sep 10. URL: https://www.brookings.edu/research/a-cure-for-health-care-inefficiency-the-value-and-geography-of-venture-capital-in-the-digital-health-sector/ [accessed 2021-08-02]
16. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851. [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]

17. World Health Organization. Report on the WHO Symposium on the Future of Digital Health Systems in the European Region. 2019 Presented at: World Health Organization Symposium; February 6-8; Copenhagen, Denmark URL: https://apps.who.int/iris/bitstream/handle/10665/329032/9789289059992-eng.pdf
18. Almathami HKY, Win KT, Vlahu-Gjorgievska E. Barriers and facilitators that influence telemedicine-based, real-time, online consultation at patients' homes: systematic literature review. J Med Internet Res 2020 Feb 20;22(2):e16407 [FREE Full text] [doi: 10.2196/16407] [Medline: 32130131]
19. Irizarry T, DeVito Dabbs A, Curran CR. Patient portals and patient engagement: a state of the science review. J Med Internet Res 2015;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
20. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
21. Damschroder LJ, Lowery JC. Evaluation of a large-scale weight management program using the consolidated framework for implementation research (CFIR). Implement Sci 2013;8:51 [FREE Full text] [doi: 10.1186/1748-5908-8-51] [Medline: 23663819]
22. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
23. Suri H. Purposeful sampling in qualitative research synthesis. Qual Res J 2011 Aug 03;11(2):63-75. [doi: 10.3316/qrj1102063]
24. 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]
25. Meskó B, Drobni Z, Bényei É, Gergely B, Győrffy Z. Digital health is a cultural transformation of traditional healthcare. Mhealth 2017;3:38 [FREE Full text] [doi: 10.21037/mhealth.2017.08.07] [Medline: 29184890]
26. Lennon MR, Bouamrane M, Devlin AM, O'Connor S, O'Donnell C, Chetty U, et al. Readiness for delivering digital health at scale: lessons from a longitudinal qualitative evaluation of a national digital health innovation program in the United Kingdom. J Med Internet Res 2017 Feb 16;19(2):e42 [FREE Full text] [doi: 10.2196/jmir.6900] [Medline: 28209558]
27. Gordon WJ, Landman A, Zhang H, Bates DW. Beyond validation: getting health apps into clinical practice. NPJ Digit Med 2020;3:14 [FREE Full text] [doi: 10.1038/s41746-019-0212-z] [Medline: 32047860]
28. Blumenthal D. Realizing the value (and profitability) of digital health data. Ann Intern Med 2017 Jun 06;166(11):842-843. [doi: 10.7326/M17-0511] [Medline: 28462419]
29. Van Winkle B, Solad Y, Vaswani N, Rosner BI. Navigating the digital health ecosystem to bridge the gap from innovation to transformation: a NODE health perspective on digital evidence. Digit Biomark 2019;3(2):83-91 [FREE Full text] [doi: 10.1159/000500194] [Medline: 32095768]
30. Muhos M, Saarela M, Foit D, Rasochova L. Management priorities of digital health service start-ups in California. Int Entrep Manag J 2018 Oct 27;15(1):43-62. [doi: 10.1007/s11365-018-0546-z]
31. Nathanson JM. Driving outcomes-digital health business models. In: Wulfovich S, Meyers A, editors. Digital Health Entrepreneurship. Cham: Springer; Jun 21, 2019:17-34.
32. Flannery D, Jarrin R. Building a regulatory and payment framework flexible enough to withstand technological progress. Health Aff (Millwood) 2018 Dec;37(12):2052-2059. [doi: 10.1377/hlthaff.2018.05151] [Medline: 30633665]
33. Singh K, Meyer SR, Westfall JM. Consumer-facing data, information, and tools: self-management of health in the digital age. Health Aff (Millwood) 2019 Mar;38(3):352-358. [doi: 10.1377/hlthaff.2018.05404] [Medline: 30830832]
34. Keesara S, Jonas A, Schulman K. Covid-19 and health care's digital revolution. N Engl J Med 2020 Jun 04;382(23):e82. [doi: 10.1056/NEJMp2005835] [Medline: 32240581]
35. Nouri S, Khoong E, Lyles C, Karliner L. Addressing equity in telemedicine for chronic disease management during the covid-19 pandemic. NEJM Catalyst. 2020 May 04. URL: https://catalyst.nejm.org/doi/abs/10.1056/CAT.20.0123 [accessed 2021-08-02]
36. Pagoto S, Bennett GG. How behavioral science can advance digital health. Transl Behav Med 2013 Sep;3(3):271-276 [FREE Full text] [doi: 10.1007/s13142-013-0234-z] [Medline: 24073178]

Abbreviations

CFIR: Consolidated Framework for Implementation Research

Edited by R Kukafka; submitted 08.10.20; peer-reviewed by L Beal, W Gordon, P Boisrond, A Campbell; comments to author 09.12.20; revised version received 19.01.21; accepted 30.06.21; published 26.08.21.

Please cite as:
Lyles CR, Adler-Milstein J, Thao C, Lisker S, Nouri S, Sarkar U
Alignment of Key Stakeholders' Priorities for Patient-Facing Tools in Digital Health: Mixed Methods Study
J Med Internet Res 2021;23(8):e24890
URL: https://www.jmir.org/2021/8/e24890
doi:10.2196/24890
PMID:34435966

©Courtney Rees Lyles, Julia Adler-Milstein, Crishyashi Thao, Sarah Lisker, Sarah Nouri, Urmimala Sarkar. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

The Effect of Collaborative Reviews of Electronic Patient-Reported Outcomes on the Congruence of Patient- and Clinician-Reported Toxicity in Cancer Patients Receiving Systemic Therapy: Prospective, Multicenter, Observational Clinical Trial

Andreas Trojan[1*], MD; Nicolas Leuthold[2*], MMed; Christoph Thomssen[3], MD; Achim Rody[4], MD; Thomas Winder[5,6], MD, PhD; Andreas Jakob[7], MD; Claudine Egger[8], MD; Ulrike Held[9], PhD; Christian Jackisch[10], MD

[1]OnkoZentrum Zürich, Zurich, Switzerland

[2]Clinic for Clinical Pharmacology and Toxicology, University Hospital Zurich, Zurich, Switzerland

[3]Universitätsklinikum Halle (Saale), Halle, Germany

[4]Klinik für Frauenheilkunde und Geburtshilfe, Universitätsklinikum Schleswig-Holstein, Lübeck, Germany

[5]Department of Internal Medicine II, Academic Teaching Hospital Feldkirch, Feldkirch, Austria

[6]University of Zurich, Zurich, Switzerland

[7]Tumor Zentrum Aarau, Hirslanden Medical Center, Aarau, Switzerland

[8]Spital Limmattal, Schlieren, Switzerland

[9]Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Zurich, Switzerland

[10]Sana Klinikum Offenbach, Offenbach, Germany

[*]these authors contributed equally

Corresponding Author:
Andreas Trojan, MD
OnkoZentrum Zürich
Seestrasse 259
Zurich, 8038
Switzerland
Phone: 41 43 344 33 33
Email: trojan@1st.ch

Abstract

Background: Electronic patient-reported outcomes (ePRO) are a relatively novel form of data and have the potential to improve clinical practice for cancer patients. In this prospective, multicenter, observational clinical trial, efforts were made to demonstrate the reliability of patient-reported symptoms.

Objective: The primary objective of this study was to assess the level of agreement κ between symptom ratings by physicians and patients via a shared review process in order to determine the future reliability and utility of self-reported electronic symptom monitoring.

Methods: Patients receiving systemic therapy in a (neo-)adjuvant or noncurative intention setting captured ePRO for 52 symptoms over an observational period of 90 days. At 3-week intervals, randomly selected symptoms were reviewed between the patient and physician for congruency on severity of the grading of adverse events according to the Common Terminology Criteria of Adverse Events (CTCAE). The patient-physician agreement for the symptom review was assessed via Cohen kappa (κ), through which the interrater reliability was calculated. Chi-square tests were used to determine whether the patient-reported outcome was different among symptoms, types of cancer, demographics, and physicians' experience.

Results: Among the 181 patients (158 women and 23 men; median age 54.4 years), there was a fair scoring agreement (κ=0.24; 95% CI 0.16-0.33) for symptoms that were entered 2 to 4 weeks before the intended review (first rating) and a moderate agreement (κ=0.41; 95% CI 0.34-0.48) for symptoms that were entered within 1 week of the intended review (second rating). However, the level of agreement increased from moderate (first rating, κ=0.43) to substantial (second rating, κ=0.68) for common symptoms of pain, fever, diarrhea, obstipation, nausea, vomiting, and stomatitis. Similar congruency levels of ratings were found for the most frequently entered symptoms (first rating: κ=0.42; second rating: κ=0.65). The symptom with the lowest agreement was hair loss (κ=–0.05). With regard to the latency of symptom entry into the review, hardly any difference was demonstrated between symptoms that were entered from days 1 to 3 and from days 4 to 7 before the intended review (κ=0.40 vs κ=0.39, respectively). In contrast, for symptoms that were entered 15 to 21 days before the intended review, no congruency was demonstrated (κ=–0.15). Congruency levels seemed to be unrelated to the type of cancer, demographics, and physicians’ review experience.

Conclusions: The shared monitoring and review of symptoms between patients and clinicians has the potential to improve the understanding of patient self-reporting. Our data indicate that the integration of ePRO into oncological clinical research and continuous clinical practice provides reliable information for self-empowerment and the timely intervention of symptoms.

Trial Registration: ClinicalTrials.gov NCT03578731; https://clinicaltrials.gov/ct2/show/NCT03578731

(J Med Internet Res 2021;23(8):e29271) doi:10.2196/29271

KEYWORDS

cancer; consilium; app; eHealth; ePRO; CTCAE; congruency; patient-reported; symptoms

Introduction

Patient-reported outcomes (PRO), such as symptoms and functional status, are commonly measured in clinical trials. There is growing interest in integrating electronic PRO (ePRO) into routine clinical practice during chemotherapeutic and immunotherapeutic interventions. Most cancer patients are motivated to spend time and effort documenting symptoms during their consultation for shared reporting with physicians. Patients’ self-empowerment and self-reporting should also improve patient-clinician communication, symptom detection, and symptom control [1]. As patient experience has gained importance in regulatory decision-making, patient-reported data are increasingly being used for quality assessment and comparative effectiveness research. Mobile health solutions have the potential to improve electronic symptom documentation, and when the collection of such PRO is widely used, it facilitates communication among stakeholders [1,2]. Several apps have been designed and tested with input from patients, nurses, and physicians. These apps have gained attention and quality with respect to improving the efficacy and safety data in oncology trials and drug discovery [3-5]. Their benefits in real-world digital patient monitoring during cancer immunotherapy have been demonstrated in terms of more accurate symptom assessment, better patient-physician communication, and reduced need for telephone consultations [6-8]. As oncologists intend to share information on symptom grading with their patients, as defined by the Common Terminology Criteria for Adverse Events (CTCAE) standards, reliable information on PRO should not only improve symptom management but also allow for the reduction of emergency admissions and improve patients’ quality of life. However, early responsiveness to symptoms and presumably longer continuation of chemotherapy, as well as a potential benefit of follow-up integration of ePRO for symptom monitoring during routine cancer care, frequently involve patient-physician or patient–nurse specialist communication [9,10]. In addition, compliance rates and the use of symptom alerts seem to be enhanced by structured graphic displays on outcome reporting [3,4]. Several digital platforms are currently implementing the capture of ePRO to allow for the sharing of data with treatment teams or to apply automatic algorithms for alert notifications in a timely and structured manner if symptoms worsen [2,11,12]. The consilium care smartphone app continuously allows oncologists to monitor the progress of patients’ symptoms through visualized progression charts based on structured patient entries. In the case of severe symptoms that exceed a determined threshold, the app notifies the patient to contact the treatment center. Previous published breast cancer studies showed the potential of the app to stabilize daily functional activity and well-being of patients in collaboration with the physician [1]. In addition, more distinct symptom entries were received from those users who shared reporting with their physicians. The functionality and utility of 2 comparable app versions for collecting ePRO have also demonstrated that the request for a collaborative review of ePRO for shared reporting increases the number of data entries and potentially affects the ability to deal with the symptoms of illness [13]. Since clinical oncology strives for a standardized recording of adverse events, the congruence between doctor and patient should serve as an important indicator that patients’ self-reporting can enhance the quality of outcome data for the accuracy of clinician ratings and safety. This has the potential to reduce the problem of patients reporting high symptom severity while their clinicians note low toxicity grades. Further, it has the potential to identify challenges in effective patient-clinician communication regarding symptom experience, to stimulate the processes of recording and reviewing patient-reported symptoms, to facilitate consultation with oncologists, and to provide self-care algorithms for real-time interventions that reduce symptom severity [13].

In this study, we evaluated the efforts being made using the consilium care app in a cohort of patients with breast, colon, lung, or prostate cancer, as well as those with hematological malignancies, to demonstrate the reliability of electronically captured patient-reported symptom entries for shared reporting with the physician to detect critical symptoms in routine cancer care. For this study, we intended to demonstrate that a collaborative review of randomly selected patient-reported symptoms improves congruency of patient- and clinician-reported toxicity in patients receiving systemic anticancer therapy. In particular, we examined whether important and frequent symptoms, such as pain, fever, diarrhea, obstipation, nausea, vomiting, and stomatitis, can be described appropriately according to the CTCAE in order to potentially implement recommendations for mitigation.

Methods

Study Design

We conducted a multicenter, observational, noninterventional study. The protocol was approved by the competent regulatory ethics committee (KEK-ZH:2017-02028) and registered on ClinicalTrials.gov (NCT03578731). Patients with breast, colon, prostate, or lung cancer, as well as those with hematological malignancies, aged 18 years and older, and initiating adjuvant or neoadjuvant systemic therapy were eligible to participate after providing written informed consent. In addition, participants had to speak German and own a smartphone. Eligible participants were recruited consecutively and without preselection according to the recommendation of the local tumor boards in centers in Switzerland, Germany, and Austria.

Objective

The primary objective of this study was to assess the level of agreement, κ, between symptom ratings by physicians at the time of the regular consultation and the ratings derived from the daily PRO between consultations. The level of agreement was analyzed in order to determine the reliability and utility of self-reported electronic symptom monitoring.

Mobile App

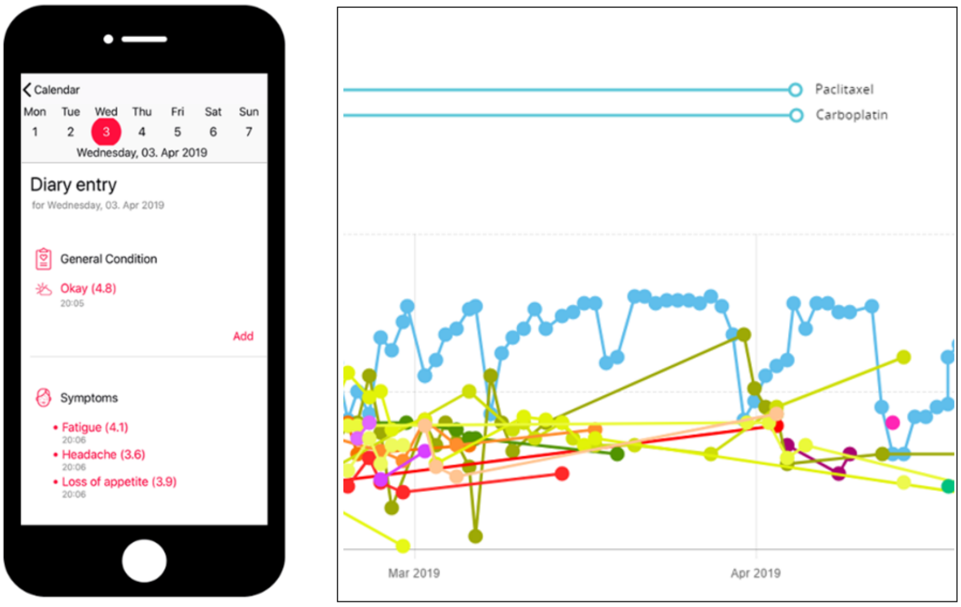
To begin, patients downloaded the consilium care app (available for iOS and Android) and connected themselves via a quick response code to their study centers. For the patients' convenience, a summary of diagnostic workup, treatment medication, and contact information of the respective treatment center was entered into the consilium care web app—the treatment team's counterpart to the smartphone app.

The app (Figure 1) facilitated the selection of well-being, symptoms, medication, and private notes. Symptoms, which were structured in groups according to organ systems, could be selected. The symptom entry display (52 distinct symptoms were available for which severity, onset, and duration could be indicated) was equipped with date and time stamps. Symptom severity, with descriptions based on the CTCAE, could be selected via a slider. The symptom history was displayed on a timeline with individual colors for each symptom. In addition, diary entries and information on diagnosis and therapy were indicated separately.

Patients were encouraged to capture data on well-being and symptoms on a daily basis. Recording usually started on the day of the therapy's initiation or the change in therapy and continued through an observational period of 12 weeks. The app allowed the continuous recording of well-being and symptoms based on the CTCAE through use of virtual analogue scales. Definitions for CTCAE grades were displayed above the slider, with which the grade of the entry could be selected via the virtual analogue scales. The severity level of a symptom, as rated by physicians and patients, was measured on an ordinal scale, with 0 indicating the lowest possible degree of severity and 4 indicating the highest possible one. The history of recorded data was displayed and visualized in the form of a symptom progression chart. In the case of severe symptoms, patients were encouraged by push notifications to seek medical advice. In addition, patients recorded their well-being according to the Eastern Cooperative Oncology Group (ECOG) performance status via a slider, with possible impairments in daily functional activities being displayed. Information for self-care (derived from the Swiss Cancer League and the Sächsische Krebsgesellschaft) was provided to them via the app depending on the severity of symptoms upon data entry.

Functional data security was ensured by identification being made only possible through the patient's ID. The data on the patient's device were encapsulated in the app, and data exchange was encrypted with the patient ID. At the study center, personal data were kept strictly separate from the data collected by the app. Data matching was performed by using the patient ID.

Figure 1. Entrance screen and a representative symptom history chart with indication of medication, well-being (blue graph), and various symptoms presented in different colors.



Collaborative Symptom Reviews

Patients were assigned to medical oncology visits every 3 weeks and invited for shared reporting and intended symptom review, which were preferably scheduled on days of therapeutic intervention. Some exceptions were made for reviews to be carried out over the phone. At the scheduled visit, the app was triggered to randomly select 2 patient-reported symptoms from the past 20 days. A first measurement of congruence (symptom 1) was performed on a symptom that was entered 2 to 3 weeks (14 to 21 days) before the actual consultation, whereas a second measurement (symptom 2) was performed on a symptom that was entered within the previous week (1 to 7 days). Patients and physicians were then prompted to perform a detailed, shared review of these symptoms in order to focus on the collection and appropriate interpretation for symptom severity grading. Up to 4 such reviews were planned per patient, including 2 electronic symptom entries per review.

Questionnaire

At the end of the observational period, participants were asked to complete a questionnaire on paper regarding the usability and usefulness of the app to clarify quality of care and the relationship between the patient and physician during the course of treatment. To this end, a 5-point Likert scale was used, with a rating from 1 (disagree) to 5 (agree very strongly).

Sample Size

We calculated the sample size on a 5% significance level to test the level of agreement, κ=0.5, between 2 raters (ie, fair to good agreement) with a precision of 0.1 on each side of 156 patients. In order to estimate κ with the necessary precision within these subgroups, we included at least 170 patients with breast cancer and 170 patients with colon cancer. We anticipated a difficultly in recruiting the same number of patients with lung cancer or prostate cancer due to their lower prevalence. Thus, the aim was to include 130 patients with either lung cancer (not fewer than 50) or prostate cancer and 130 patients with hematological malignancies. We planned to enroll a total of 600 patients, as we expected 15% to 20% of enrolled patients to discontinue participation (dropout) early.

The originally planned study population size for the entire study cohort was 600. The study duration was estimated to be about 3 years, starting in March 2018. In autumn 2020, only about one-third of the planned study patients were recruited, and the sponsor decided to prematurely terminate the study on October 11, 2020, due to insufficient recruitment. Despite the continuous opening of many study sites beginning 2018, due to the present recruitment rate and the ongoing COVID-19 situation, the planned number of 600 patients was unachievable.

Statistical Analysis

Descriptive statistics included mean and SD for continuous variables, and numbers and percentages of total for categorical variables. For statistical analysis, the associations between physicians' and patients' ratings were visualized by plots. Multiple ratings for patients were included and accounted for by the analysis. For the quantification of levels of agreement, Cohen kappa (κ) values were calculated with squared weights. κ values are reported with 95% CIs. These CIs were based on 1000 bootstrap samples. According to Landis and Koch [14], values for κ were characterized as follows: <0, no agreement; 0 to 0.20, slight agreement; 0.21 to 0.40, fair agreement; 0.41 to 0.60, moderate agreement; 0.61 to 0.80, substantial agreement; and 0.81 to 1, almost perfect agreement. All analyses were carried out with R version 4.0.2 (The R Foundation for Statistical Computing) [15], and Excel R Markdown was used for dynamic reporting.

Results

Baseline Characteristics

Between February 2018 and October 2020, 223 patients (190 female and 33 male) with cancer (170 breast, 19 lung, 15 colon, 7 prostate, and 12 hematological [B cell] malignancies) were included using the consilium care app. Among them, 181 patients (158 women and 23 men; age at therapy start: mean 54.4 years, SD 12.1) had performed at least 1 validated review with the treating physician. About half of the 181 patients who used the consilium care app were treated in an adjuvant setting (vs neoadjuvant). Fewer than one-third (51/181, 28.2%) of the patients received treatment for advanced disease with noncurative intention. In total, 27 distinct chemotherapeutic agents in 17 different chemotherapy regimens were administered, including antihormones, CDK4/6 inhibitors, and immunotherapies.

Due to the lack of appropriate accrual within the context of the COVID-19 pandemic, premature closing of the study, and other issues, 42 patients included could not perform a minimum of 1 intended review. In addition to this, 7 patients were not evaluable due to the premature study termination, 10 patients did not enter a sufficient number of symptoms, and another 14 patients were not evaluable due to technical issues. Only 3 patients withdrew their informed consent. Baseline characteristics are displayed in Table 1, and an overview flow chart of the patient enrollment is available in Multimedia Appendix 1.

Table 1. Baseline characteristics.

Characteristic	Value (N=181)
Primary tumor, n (%)	
Hematological	9 (5.0)
Breast	142 (78.5)
Colon	11 (6.1)
Lung	13 (7.2)
Prostate	6 (3.3)
Sex, n (%)	
Female	157 (86.7)
Male	23 (12.7)
N/A[a]	1 (0.6)
Age at start, mean (SD)	54.4 (12.1)

[a] N/A: not applicable.

Agreement Levels

A total of 181 patients underwent at least 1 intended symptom review for this analysis. From a subset of 110 patients (60.8%), more than 2 collaborative symptom reviews of patients with their physicians were available for analysis. For the analysis of the first symptom agreement levels (across all multiple ratings per patient), there were 497 (first rating) reviews available for analysis, while for the second symptom agreement levels, 483 reviews (second rating) were available.

An estimation of general agreement levels between physicians' and patients' observations in the first symptom (defined as recorded 14 to 21 days before the review) revealed a fair congruency of κ=0.24 (95% CI 0.16-0.33), while for the second most recent symptom (defined as being recorded 1 to 7 days before), the value rose to κ=0.41 (95% CI 0.34-0.48; Figure 2).

Analysis of the levels of agreement in subgroups of the specific symptoms, including pain, fever, diarrhea, obstipation, nausea, vomiting, and stomatitis, revealed a higher congruency between the patient and physician estimate (symptom 1: κ=0.43, 95% CI 0.21-0.62; symptom 2: κ =0.68, 95% CI 0.54-0.77; Figure 3). Whether this observation was due to a different perception of clinical relevance and frequency of these symptoms or to a clearer description, as 5 of the 7 symptoms were associated with objectifiable values in their definition (eg, fewer than 4 loose stools per day) at some point, remains unclear.

Next, we evaluated the levels of agreement in the subgroup of physicians with at least 10 ratings. The distribution of rating frequencies revealed large differences; of the 29 participating in this study, 9 physicians performed 10 or more ratings. These were considered experienced raters and were included in the subsequent assessments. For the analysis, there were 417 observations for symptom 1 (first rating) and 405 observations for symptom 2 (second rating). Again, multiple ratings per patient were included. As shown in Figure 4, a fair congruency between patient and physician estimates was present for those considered experienced (≥10 ratings; symptom 1: κ =0.25, 95% CI 0.17-0.34; symptom 2: κ=0.41, 95% CI 0.33-0.49). Compared to all physicians' (experienced and less experienced) ratings for symptom 1 (κ=0.24) and symptom 2 (κ=0.41), the agreement levels hardly differed, indicating that congruency was more likely affected by timing and symptom description than the physicians' particular skills.

Similar results of congruency as those seen in the specific symptoms displayed in Figure 2 were obtained for the most frequent symptom as rated by experienced physicians (>10 ratings; symptom 1: κ=0.42, 95% CI 0.18-0.62; symptom 2: κ =0.65, 95% CI 0.5-0.75; Figure 5). The most frequently captured symptoms were fatigue, hot flashes, sleep disorder, headache, and taste disorder.

The levels of agreement with respect to time intervals between the date of collaborative review and the date of symptom entry within the previous week did not reveal a significant difference (days 1-3: κ=0.40; days 4-7: κ=0.39; overall days 1-7: κ=0.41). For the rating of symptoms entered 15 to 21 days prior to the review, a significant lack of congruency was noted (κ=–0.15). This finding indicated that patients recalled symptoms and their severity much better if they occurred more recently. For future studies, a collaborative review of a symptom from the recent past may be considered sufficient to demonstrate the accuracy of the electronic symptom recording in general, particularly for distinct and frequently occurring symptoms. Although this observation might require confirmation in a subsequent study, the idea of recent-past symptom validation (less than 7 days) might be applicable in real-world cancer care, clinical trials, or pay-for-performance models [16]. Furthermore, we noted a moderate increase of congruence between ratings from week 3 (first rating) to week 9 (third rating) in our approach (symptom 1: κ=0.23 vs κ=0.29; symptom 2: κ=0.36 vs κ=0.41), indicating a potential training effect in patients and physicians. The quality of ratings neither appeared differently with regard to light or moderate symptoms (CTCAE grade ≤ 2) nor in comparison to severe symptoms, defined as CTCAE grade >2 (κ=0.13 vs κ=0.11), which is important in cases of early-intervention clinical practice. Congruency of symptom reporting according to the review of the second symptom was similar for breast (396 reviews; κ=0.39), lung (30 reviews; κ=0.45), and colon cancer (23 reviews; κ=0.51), as well as hematological malignancies (20 reviews; κ=0.49). For prostate cancer, there was an almost perfect congruency (12 reviews; κ=0.82) although the low number of reviews had to be considered with regard to statistical significance. The subgroup analysis for age and gender showed overall congruency levels of κ=0.50 for older (>65 years; 99 reviews; κ=0.50;) and younger patients (<65 years; 380 reviews; κ=0.38), as well as for female (435 reviews; κ=0.40) and male (44 reviews; κ=0.49 for) patients.

Figure 2. Estimations of agreement levels between physicians' and patients' observations for the first and second symptom. diarr: diarrhea; fev: fever; obstip: obstipation; stomat: stomatatis; vomit: vomiting.

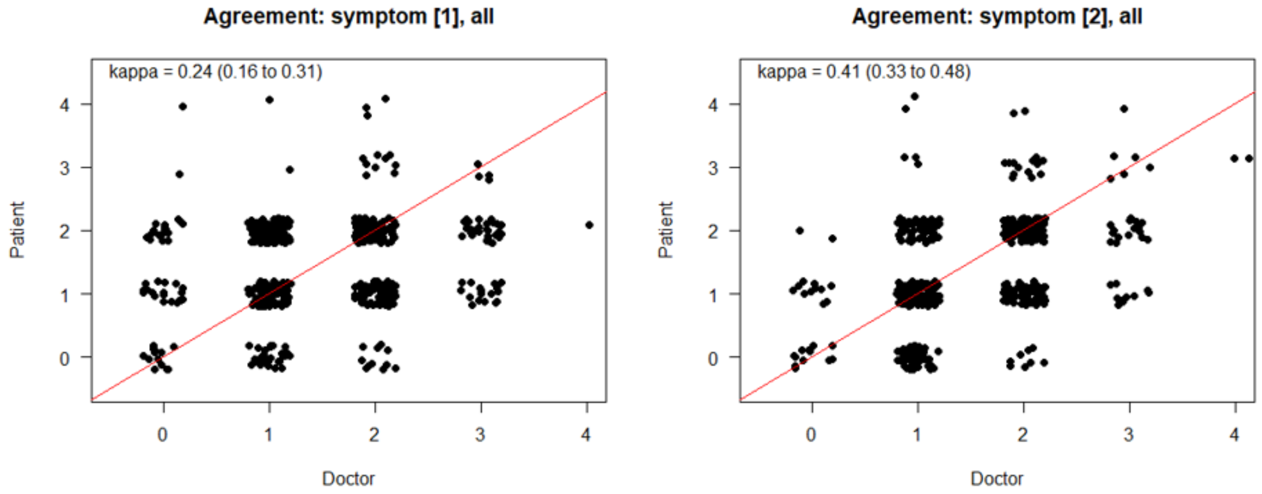


Figure 3. Estimations of agreement levels between physicians' and patients' observations for specific symptoms.

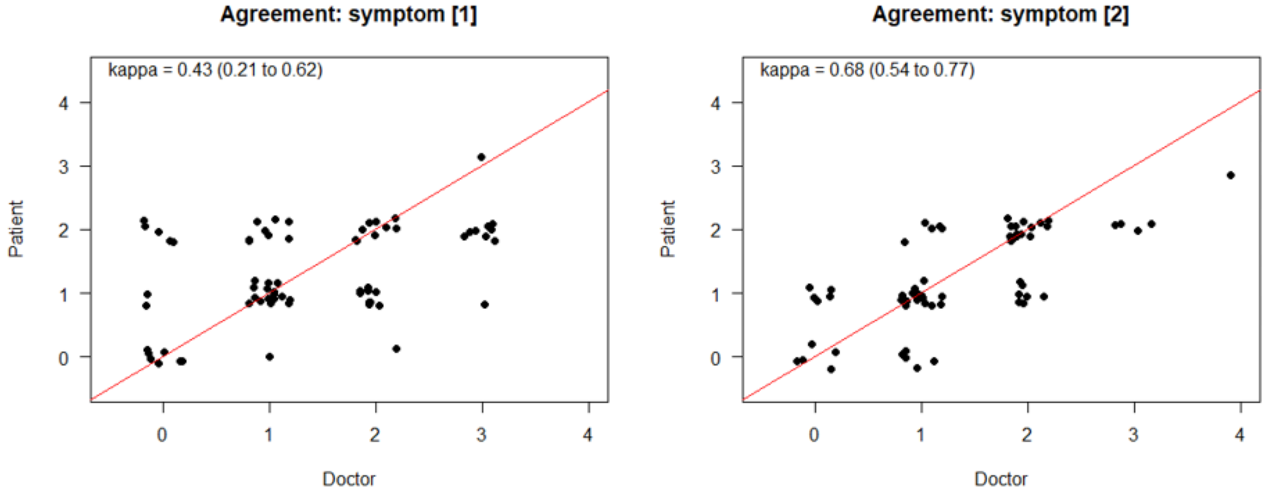


Figure 4. Estimations of agreement levels between physicians' and patients' observations for experienced physicians.

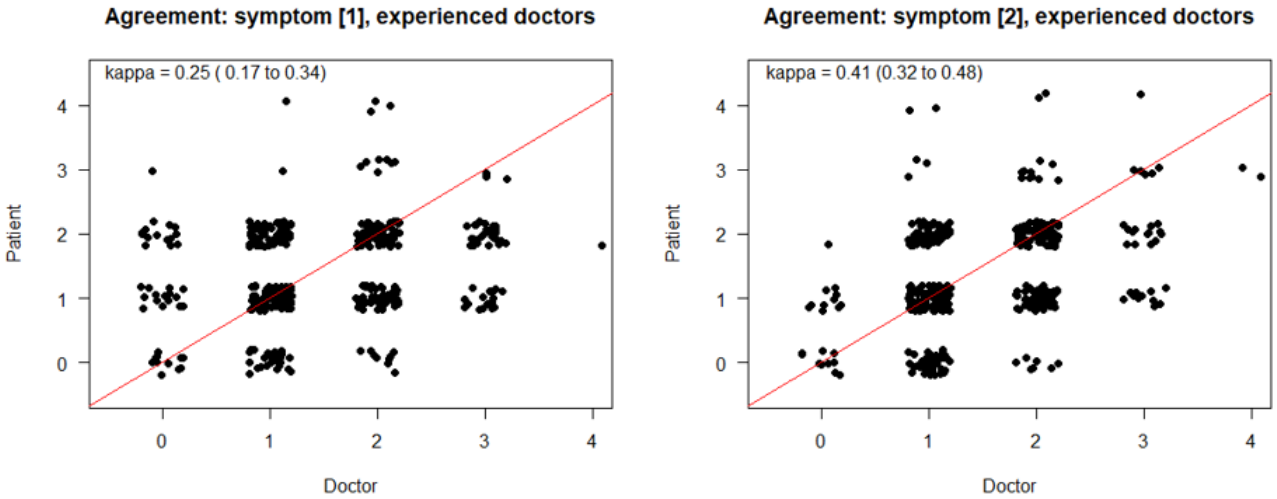
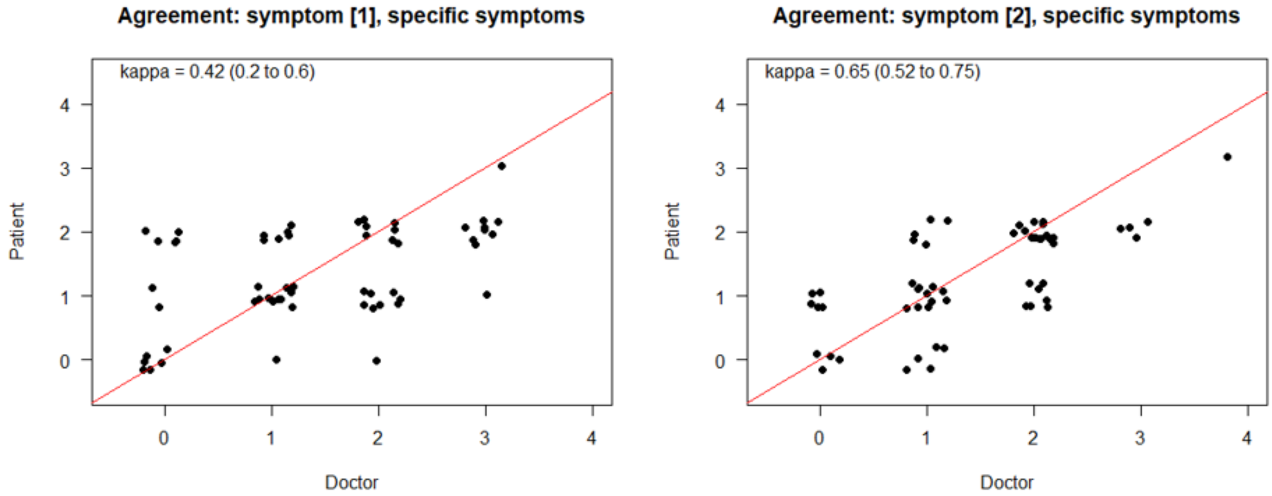


Figure 5. Estimations of agreement levels of experienced physicians specifically for the most frequent symptoms.

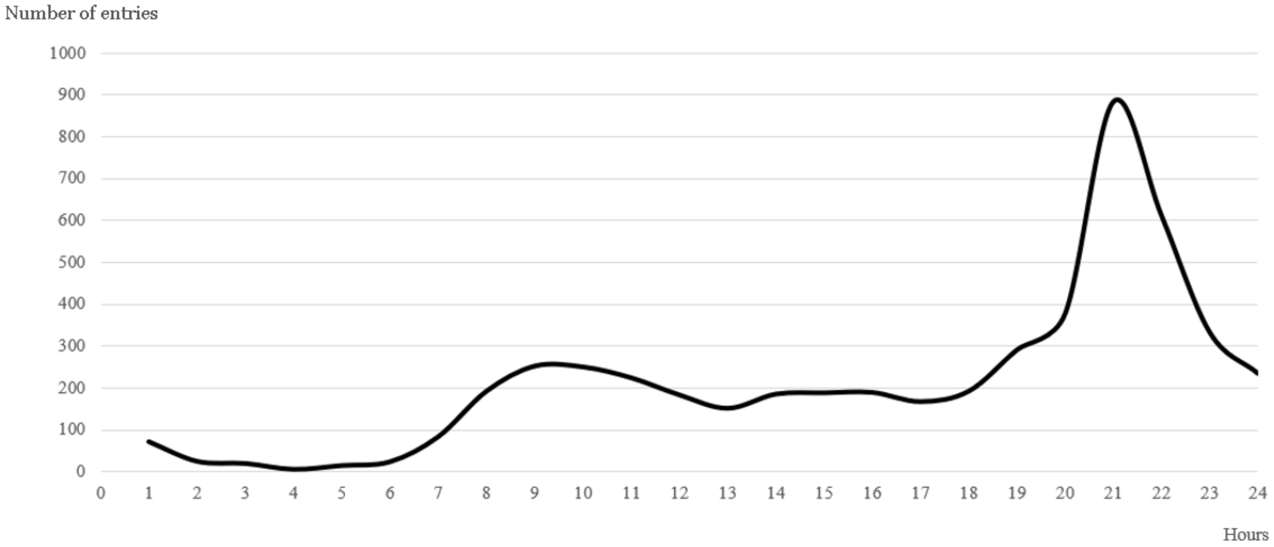


Well-being and Symptoms

Regarding well-being, 4762 data entries were derived from 210 evaluable patients during the observation period. Patients reported their well-being almost every single day and in a classical circadian rhythm (Figure 6). Because well-being was reported independently of the underlying diagnosis or symptoms, we assumed that this indicated a pattern of app use. Users preferred to use the app in the morning and also used it during the evening hours. Therefore, a circadian pattern of symptom reporting seemed to be favored. The degree to which the app's functions (eg, occasional push notifications, design features, tips for self-care, or effects of collaborative review and shared reporting) affected data entries remains unclear, as this evaluation was not addressed.

Overall, 210 patients generated a large absolute number of 42,142 electronically reported symptoms and side effects, suggesting easy handling of the app for an effective symptom history insight. Given the observational period of 84 days, this resulted in an average number of 2 to 3 entries per patient and day. The most commonly reported symptom was fatigue, which was indicated significantly more often in the breast cancer and lymphoma groups (data not shown) compared to other cancer entities. Due to the heterogeneity of drugs and limited information on dosage, a potential association of symptoms with the respective cancer type, medication, or regimen, could not be performed sufficiently. However, more than 32.59% (13,734/42,142) of all data entries affected usual activities of daily living and symptoms such as pain/discomfort (8370/42,142, 19.86%), self-care (3475/42,142, 8.24%), anxiety/depression (1458/42,142, 3.45%), and mobility (431/42,142, 1.02%), all of which potentially represent components of the 5-level EQ-5D questionnaire.

Figure 6. Circadian distribution of well-being entries (y-axis) over 24 hours (x-axis).



Unplanned Consultations and Serious Adverse Events

Although fewer than 18.2% (33/181) of the participants with solid cancer (breast, colon, lung, prostate) required unplanned consultations or emergency services due to treatment-related side effects and toxicities, more than twice this proportion (4/9, 44%) was recorded in patients with lymphoma, mostly attributed due to fatigue and fever. An association with a possible benefit from app use cannot be made, as data from a matched analysis (age, cancer type, therapy) of patients from 2 larger participating cancer centers indicated only a nonsignificant decline in these events (data not shown). Importantly, no serious adverse events related to the use of the app were recorded during the entire study period.

Usability and Usefulness of the App

Questionnaires from 171 patients included were available for the rating of the app at the time of this survey. Six patients died due to cancer progression during the study, from whom surveys were not available for analysis. A utility analysis could not be conducted on 16 patients, as they were not correctly included into the study, withdrew informed consent, had technical problems, or lacked a sufficient number of data entries. The results are displayed in Multimedia Appendix 2.

Discussion

The systematic electronic recording of PRO by smartphone has not yet been extensively explored in cancer treatment. Previous studies indicate that the range of measures used and symptoms captured seem to vary greatly across studies, and that, regardless of the concordance metric employed, the reported agreement between clinician-based CTCAE and PRO seems to be moderate, at best [17]. In one study that retrospectively applied CTCAE patient language adaptations, including the Symptom Tracking and Reporting system, to assess specific symptoms, extracted clinician- and patient-reported adverse event ratings were considered poor to moderate, at best, when the applied rating sources for each of the adverse events were compared [18]. In an attempt to improve these differences, we explored integrating ePRO and clinician reporting with a standardized, shared review process, according to CTCAE criteria with adapted patient-oriented language by testing the level of agreement between the patients' and physicians' judgment on the severity of patient symptoms with 3 weekly reviews of randomly selected symptoms at any severity grade.

Overall, we found fair agreement for long-lasting symptoms, whereas for the more recent symptoms (defined as those recorded 1-7 days earlier), the degree of agreement in symptom reporting between the patient and physician was moderate and comparable to results from a study in early breast cancer [19]. However, the congruence between patients and physicians gained substantial reliability when analyses on levels of agreement in subgroups of the specific symptoms (ie, pain, fever, diarrhea, obstipation, nausea, vomiting, and stomatitis) were performed and also in an identical manner to that in the most frequently occurring symptoms, including fatigue, hot flashes, sleep disorder, headache, and taste disorder. Together, data entries from these symptoms covered about 50% of all recorded symptom-related entries during this study. As patients obviously recalled recent symptoms more clearly, the high trustworthiness of symptom rating could be sufficiently proven by 1 review in this context. Congruency of rating seemed to be independent of the reviewers' experience, and no outlier result in congruency of symptom reporting could be demonstrated for any specific patient cohort, indicating the potentially broad acceptance and use of such an approach. Additionally, no differences in symptom congruency were noted with respect to light or more severe symptom grading.

Compliance for the use of the consilium care app was high as evidenced by the high number of 2 to 3 data entries per patient and by the response from questionnaires, and was found to be comparable with results from other studies that used more standardized questionnaires for different devices [20]. In a recent study, patients were invited to complete the European Organization for Research and Treatment of Cancer (EORTC) Core Quality of Life questionnaire (QLQ-C30) and cancer site–specific modules before each visit on tablets or computers in the hospital or at home. An adequate compliance (at least 66% of health-related quality of life assessments were completed) was demonstrated for the cohort of breast cancer (96%), colorectal cancer (98%), and lung cancer (91%) [21], which we consider comparable to the results of our study.

In one study that administered weekly PRO from the National Cancer Institute's PRO-CTCAE item library (for symptoms such as pain, nausea, and diarrhea) via mail or telephone and assessed them by using a 5-point ordinal verbal descriptor scale and via PRO questions about physical performance (ECOG) and financial toxicities (The Comprehensive Score for Financial Toxicity [COST-FACIT] questionnaire), it was found that most patients agreed that weekly reporting was a favored frequency for ePRO questionnaire administration in the context of advanced and metastatic cancer treatment [22]. However, during a more complex or intensive treatment phase, a more frequent (even daily) assessment of more than 8 symptoms might be well regarded and positively associated with an increased use of educational materials about home symptom management. In another trial, almost 40% of patients (particularly older patients and those living in rural areas) chose to use an automated telephone interface rather than a web interface or preferred personal contact in the case of severe symptoms affecting cognitive or sexual dysfunction [23]. Web-based, guided, self-help interventions can provide clinically meaningful improvements in quality of life; however, producing a meaningful effect might require punctual psychological interventions [24]. Although no such findings were apparent in our study, following advice and using tips for timely self-care and compliance remains challenging for patients and caregivers. The consilium app contains 20 tips for the most common symptoms. In personal communication with patients, it was suggested that this opportunity of self-help intervention should be linked to the appropriate symptom or grade, as patients perceived this to be a component of personalized medicine [25,26].

There were potential limitations to this study. The frequency of the completed symptom reviews varied between the 3 German-speaking countries conducting the trial, most patients were suffering from breast cancer, and the study was not randomized, which precluded analysis in regard to the effects of empowered self-care and the potential impact on unplanned consultations. Statistical limitations evolved from the data set when there were multiple observations per patient; thus, observations could not be considered independent. Furthermore, there were limitations to the interpretation of Cohen κ values. In this study, we used magnitude guidelines proposed by Landis and Koch [14] to describe levels of interrater reliability; however, other guidelines exist, such as those of Fleiss [27]. Because of the ongoing debate about the correct description of κ values, the interpretation we employed can still be subject to scrutiny. Importantly, due to the lack of appropriate accrual in the context of the COVID-19 pandemic, the trial was ended prematurely.

In summary, we demonstrated that a shared monitoring and review process to assess symptoms between patients and physicians has the potential to improve the quality of future patient self-reporting. Our study indicated that the integration of ePRO into oncological clinical research and continuous clinical practice should leverage monitoring of side effects and symptom management [28,29] using the rapidly developing digital mobile and sensor technologies, which can provide more objective measures and facilitate the active and passive collection of detailed, personalized data.

Acknowledgments

The authors would like to thank all patients who participated in this study. Furthermore, we would like to thank the Swiss Tumor Institute, Zürich, and the Hirslanden Forschungsstiftung, Zürich, for their financial support.

Authors' Contributions

AT, UH, and CJ were responsible for the integrity of the entire study. AT, TW, and UH were responsible for the study concept and design. AT, TW, and NL conducted the literature research. AT, CT, AR, TW, AJ, CE, and NL conducted the trial. UH, NL, and AT performed the data analysis. UH and NL performed statistical analyses. NL, AT, UH, and CJ prepared the manuscript. NL, AT, UH, and CJ edited the manuscript. NL and AT contributed equally as the main authors. All authors approved the final manuscript.

Conflicts of Interest

AT is the founder and chief medical officer of Mobile Health AG, a startup company that operates the consilium care smartphone app. He also owns stock in the company. The other authors have no conflicts of interest to declare.

Multimedia Appendix 1
Overview flow chart of the patient enrollment.
[PNG File , 37 KB - jmir_v23i8e29271_app1.png]

Multimedia Appendix 2
Results of the questionnaire about the usability and usefulness of the app.
[PNG File , 57 KB - jmir_v23i8e29271_app2.png]

References

1. Egbring M, Far E, Roos M, Dietrich M, Brauchbar M, Kullak-Ublick GA, et al. A mobile app to stabilize daily functional activity of breast cancer patients in collaboration with the physician: a randomized controlled clinical trial. J Med Internet Res 2016 Sep 06;18(9):e238 [FREE Full text] [doi: 10.2196/jmir.6414] [Medline: 27601354]
2. Basch E, Deal AM, Kris MG, Scher HI, Hudis CA, Sabbatini P, et al. Symptom monitoring with patient-reported outcomes during routine cancer treatment: a randomized controlled trial. J Clin Oncol 2016 Feb 20;34(6):557-565 [FREE Full text] [doi: 10.1200/JCO.2015.63.0830] [Medline: 26644527]
3. Basch E. High compliance rates with patient-reported outcomes in oncology trials submitted to the US Food and Drug Administration. J Natl Cancer Inst 2019 May 01;111(5):437-439. [doi: 10.1093/jnci/djy183] [Medline: 30561701]
4. Snyder C, Smith K, Holzner B, Rivera YM, Bantug E, Brundage M, PRO Data Presentation Delphi Panel. Making a picture worth a thousand numbers: recommendations for graphically displaying patient-reported outcomes data. Qual Life Res 2019 Feb;28(2):345-356 [FREE Full text] [doi: 10.1007/s11136-018-2020-3] [Medline: 30306533]
5. Basch E, Dueck AC. Patient-reported outcome measurement in drug discovery: a tool to improve accuracy and completeness of efficacy and safety data. Expert Opin Drug Discov 2016 Aug;11(8):753-758. [doi: 10.1080/17460441.2016.1193148] [Medline: 27310432]
6. Iivanainen S, Alanko T, Peltola K, Konkola T, Ekström J, Virtanen H, et al. ePROs in the follow-up of cancer patients treated with immune checkpoint inhibitors: a retrospective study. J Cancer Res Clin Oncol 2019 Mar;145(3):765-774 [FREE Full text] [doi: 10.1007/s00432-018-02835-6] [Medline: 30666409]
7. Trojan A, Huber U, Brauchbar M, Petrausch U. Consilium smartphone app for real-world electronically captured patient-reported outcome monitoring in cancer patients undergoing anti-PD-L1-directed treatment. Case Rep Oncol 2020;13(2):491-496 [FREE Full text] [doi: 10.1159/000507345] [Medline: 32518544]
8. Schmalz O, Jacob C, Ammann J, Liss B, Iivanainen S, Kammermann M, et al. Digital monitoring and management of patients with advanced or metastatic non-small cell lung cancer treated with cancer immunotherapy and its impact on quality of clinical care: interview and survey study among health care professionals and patients. J Med Internet Res 2020 Dec 21;22(12):e18655 [FREE Full text] [doi: 10.2196/18655] [Medline: 33346738]
9. 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]
10. Lizée T, Basch E, Trémolières P, Voog E, Domont J, Peyraga G, et al. Cost-effectiveness of web-based patient-reported outcome surveillance in patients with lung cancer. J Thorac Oncol 2019 Jun;14(6):1012-1020. [doi: 10.1016/j.jtho.2019.02.005] [Medline: 30776447]
11. Licqurish SM, Cook OY, Pattuwage LP, Saunders C, Jefford M, Koczwara B, et al. Tools to facilitate communication during physician-patient consultations in cancer care: an overview of systematic reviews. CA Cancer J Clin 2019 Nov;69(6):497-520 [FREE Full text] [doi: 10.3322/caac.21573] [Medline: 31339560]
12. Calvert M, Kyte D, Mercieca-Bebber R, Slade A, Chan A, King MT, the SPIRIT-PRO Group, et al. Guidelines for inclusion of patient-reported outcomes in clinical trial protocols: the spirit-pro extension. JAMA 2018 Feb 06;319(5):483-494. [doi: 10.1001/jama.2017.21903] [Medline: 29411037]
13. Trojan A, Bättig B, Mannhart M, Seifert B, Brauchbar MN, Egbring M. Effect of collaborative review of electronic patient-reported outcomes for shared reporting in breast cancer patients: descriptive comparative study. JMIR Cancer 2021 Mar 17;7(1):e26950 [FREE Full text] [doi: 10.2196/26950] [Medline: 33729162]
14. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [Medline: 843571]
15. The R Project for Statistical Computing. R Foundation for Statistical Computing. URL: https://www.R-project.org/ [accessed 2021-03-29]
16. Picecchi D, Bertram K, Brücher D, Bauer M. Towards novel reimbursement models for expensive advanced therapy medicinal products (ATMPs). Swiss Med Wkly 2020 Dec 14;150:w20355 [FREE Full text] [doi: 10.4414/smw.2020.20355] [Medline: 33378542]
17. Atkinson TM, Ryan SJ, Bennett AV, Stover AM, Saracino RM, Rogak LJ, et al. The association between clinician-based common terminology criteria for adverse events (CTCAE) and patient-reported outcomes (PRO): a systematic review. Support Care Cancer 2016 Aug;24(8):3669-3676 [FREE Full text] [doi: 10.1007/s00520-016-3297-9] [Medline: 27260018]
18. Atkinson TM, Rogak LJ, Heon N, Ryan SJ, Shaw M, Stark LP, et al. Exploring differences in adverse symptom event grading thresholds between clinicians and patients in the clinical trial setting. J Cancer Res Clin Oncol 2017 Apr;143(4):735-743 [FREE Full text] [doi: 10.1007/s00432-016-2335-9] [Medline: 28093637]
19. Nyrop KA, Deal AM, Reeve BB, Basch E, Chen YT, Park JH, et al. Congruence of patient- and clinician-reported toxicity in women receiving chemotherapy for early breast cancer. Cancer 2020 Jul 01;126(13):3084-3093 [FREE Full text] [doi: 10.1002/cncr.32898] [Medline: 32315091]
20. Howell D, Rosberger Z, Mayer C, Faria R, Hamel M, Snider A, iPEHOC Collaborative Team. Personalized symptom management: a quality improvement collaborative for implementation of patient reported outcomes (PROs) in 'real-world'

oncology multisite practices. J Patient Rep Outcomes 2020 Jun 17;4(1):47 [FREE Full text] [doi: 10.1186/s41687-020-00212-x] [Medline: 32556794]

21. Mouillet G, Falcoz A, Fritzsch J, Almotlak H, Jacoulet P, Pivot X, et al. Feasibility of health-related quality of life (HRQoL) assessment for cancer patients using electronic patient-reported outcome (ePRO) in daily clinical practice. Qual Life Res 2021 Jan 02. [doi: 10.1007/s11136-020-02721-0] [Medline: 33387291]
22. Basch E, Stover AM, Schrag D, Chung A, Jansen J, Henson S, et al. Clinical utility and user perceptions of a digital system for electronic patient-reported symptom monitoring during routine cancer care: findings from the PRO-TECT trial. JCO Clin Cancer Inform 2020 Oct;4:947-957 [FREE Full text] [doi: 10.1200/CCI.20.00081] [Medline: 33112661]
23. Gressel GM, Dioun SM, Richley M, Lounsbury DW, Rapkin BD, Isani S, et al. Utilizing the Patient Reported Outcomes Measurement Information System (PROMIS®) to increase referral to ancillary support services for severely symptomatic patients with gynecologic cancer. Gynecol Oncol 2019 Mar;152(3):509-513. [doi: 10.1016/j.ygyno.2018.10.042] [Medline: 30876496]
24. Urech C, Grossert A, Alder J, Scherer S, Handschin B, Kasenda B, et al. Web-based stress management for newly diagnosed patients with cancer (STREAM): a randomized, wait-list controlled intervention study. J Clin Oncol 2018 Mar 10;36(8):780-788 [FREE Full text] [doi: 10.1200/JCO.2017.74.8491] [Medline: 29369731]
25. Pircher M, Winder T, Trojan A. Response to vemurafenib in metastatic triple-negative breast cancer harbouring a BRAF V600E mutation: a case report and electronically captured patient-reported outcome. Case Rep Oncol 2021;14(1):616-621 [FREE Full text] [doi: 10.1159/000513905] [Medline: 33976643]
26. Richardson D, Zhan L, Mahtani R, McRoy L, Mitra D, Reynolds M, et al. A prospective observational study of patient-reported functioning and quality of life in advanced and metastatic breast cancer utilizing a novel mobile application. Breast Cancer Res Treat 2021 May;187(1):113-124 [FREE Full text] [doi: 10.1007/s10549-020-06082-7] [Medline: 33428072]
27. Fleiss J. Statistical Methods for Rates and Proportions (Wiley Series in Probability and Statistics). New York, USA: Wiley-Interscience; 1981.
28. Thong MSY, Chan RJ, van den Hurk C, Fessele K, Tan W, Poprawski D, MASCC EMA working group. Going beyond (electronic) patient-reported outcomes: harnessing the benefits of smart technology and ecological momentary assessment in cancer survivorship research. Support Care Cancer 2021 Jan;29(1):7-10 [FREE Full text] [doi: 10.1007/s00520-020-05648-x] [Medline: 32844316]
29. Riis CL, Stie M, Bechmann T, Jensen PT, Coulter A, Möller S, et al. ePRO-based individual follow-up care for women treated for early breast cancer: impact on service use and workflows. J Cancer Surviv 2021 Jan 07. [doi: 10.1007/s11764-020-00942-3] [Medline: 33415653]

Abbreviations

COST-FACIT: The Comprehensive Score for Financial Toxicity
CTCAE: Common Terminology Criteria of Adverse Events
ECOG: Eastern Cooperative Oncology Group
ePRO: electronic patient-reported outcomes
EORTC: European Organization for Research and Treatment of Cancer
PRO: patient-reported outcomes

Edited by G Eysenbach; submitted 05.04.21; peer-reviewed by B Koczwara, C Jacob, P Innominato; comments to author 26.04.21; revised version received 13.05.21; accepted 03.06.21; published 05.08.21.

Please cite as:
Trojan A, Leuthold N, Thomssen C, Rody A, Winder T, Jakob A, Egger C, Held U, Jackisch C
The Effect of Collaborative Reviews of Electronic Patient-Reported Outcomes on the Congruence of Patient- and Clinician-Reported Toxicity in Cancer Patients Receiving Systemic Therapy: Prospective, Multicenter, Observational Clinical Trial
J Med Internet Res 2021;23(8):e29271
URL: https://www.jmir.org/2021/8/e29271
doi:10.2196/29271
PMID:34383675

©Andreas Trojan, Nicolas Leuthold, Christoph Thomssen, Achim Rody, Thomas Winder, Andreas Jakob, Claudine Egger, Ulrike Held, Christian Jackisch. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Review

Measurement Properties of Patient-Reported Outcome Measures for Diabetes: Systematic Review

Priscilla Jia Ling Wee[1*], LLB; Yu Heng Kwan[2,3*], BSc, MD, PhD; Dionne Hui Fang Loh[4], BEng; Jie Kie Phang[5], BSc; Troy H Puar[6], MBBS, MRCP; Truls Østbye[2], MD, MPH, PhD; Julian Thumboo[2,5,7], MBBS, MMed, MRCP; Sungwon Yoon[2], MPH, PhD; Lian Leng Low[4,8,9], MBBS, MMed, MCI

[1]Duke-NUS Medical School, Singapore, Singapore
[2]Programme in Health Services and Systems Research, Duke-NUS Medical School, Singapore, Singapore
[3]Department of Pharmacy, National University of Singapore, Singapore, Singapore
[4]SingHealth Office of Regional Health, Singapore, Singapore
[5]Department of Rheumatology and Immunology, Singapore General Hospital, Singapore, Singapore
[6]Department of Endocrinology, Changi General Hospital, Singapore, Singapore
[7]Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
[8]Department of Family Medicine and Continuing Care, Singapore General Hospital, Singapore, Singapore
[9]Post Acute and Continuing Care, Outram Community Hospital, SingHealth Community Hospitals, Singapore, Singapore
[*]these authors contributed equally

Corresponding Author:
Yu Heng Kwan, BSc, MD, PhD
Programme in Health Services and Systems Research
Duke-NUS Medical School
8 College Road
Singapore, 169857
Singapore
Phone: 65 90231226
Fax: 65 65348632
Email: phakyh@nus.edu.sg

Abstract

Background: The management of diabetes is complex. There is growing recognition of the use of patient-reported outcome measures (PROMs) as a standardized method of obtaining an outlook on patients' functional status and well-being. However, no systematic reviews have summarized the studies that investigate the measurement properties of diabetes PROMs.

Objective: Our aims were to conduct a systematic review of studies investigating the measurement properties of diabetes PROMs by evaluating the methodological quality and overall level of evidence of these PROMs and to categorize them based on the outcome measures assessed.

Methods: This study was guided by the PRISMA (Preferred Reporting Items for Systematic Review and Meta-Analysis) guidelines. Relevant articles were retrieved from the Embase, PubMed, and PsychINFO databases. The PROMs were evaluated with the COSMIN (COnsensus-based Standards for the selection of health Measurement Instruments) guidelines.

Results: A total of 363 articles evaluating the measurement properties of PROMs for diabetes in the adult population were identified, of which 238 unique PROMs from 248 studies reported in 209 articles were validated in the type 2 diabetes population. PROMs with at least a moderate level of evidence for ≥5 of 9 measurement properties include the Chinese version of the Personal Diabetes Questionnaire (C-PDQ), Diabetes Self-Management Instrument Short Form (DSMI-20), and Insulin Treatment Appraisal Scale in Hong Kong primary care patients (C-ITAS-HK), of which the C-PDQ has a "sufficient (+)" rating for >4 measurement properties. A total of 43 PROMs meet the COSMIN guidelines for recommendation for use.

Conclusions: This study identified and synthesized evidence for the measurement properties of 238 unique PROMs for patients with type 2 diabetes and categorized the PROMs according to their outcome measures. These findings may assist clinicians and researchers in selecting appropriate high-quality PROMs for clinical practice and research.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42020180978; https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020180978.

(J Med Internet Res 2021;23(8):e25002) doi:10.2196/25002

KEYWORDS

systematic review; measurement properties; patient-reported outcome measures; methodological quality; level of evidence; PROMs; patient reported outcome; diabetes

Introduction

Diabetes is a serious and common chronic condition that affects approximately 425 million people worldwide between the ages of 20 and 79 years [1]. The management of diabetes is complex and multifaceted, as the disease is associated with various complications and imposes significant psychological and emotional burdens on the individual [2]. Successful diabetes care requires a systematic approach to support patients' behavior change efforts, including healthy lifestyle choices, self-management, and identification of self-management problems [2]. Hence, clinical decisions made in the management of diabetes should be patient-centered, as this approach can help clinical providers identify barriers to adherence as well as motivations for self-care [2].

Recognition is growing of the usefulness of patient-reported outcome measures (PROMs) in patient-centered care and clinical decision-making [3]. PROMs are direct reports of a patient's health status and well-being from their own perspective [4]. By providing a platform for patients to convey their disease experience, and by serving as a screening tool for underlying mental and functional problems, PROMs can bridge the gap between clinical concerns and patient perspectives, providing a more holistic assessment for enhancing diabetes care [4].

To fill this need, a considerable number of different PROMs for patients with diabetes in the adult population (>18 years of age) have been developed and revised over the last two decades. Examples include the 39-item Diabetes-39 (measuring quality of life of people with diabetes) [5] and the 20-item Problem Areas in Diabetes (PAID) Scale (measuring emotional functioning in diabetes) [6], which has since been revised to a short form 5-item scale (ie, the PAID-5 [7]). The large number of available PROMs creates challenges for clinicians or researchers to select the most appropriate high-quality PROM for their specific needs. To date, no systematic review has summarized PROMs for diabetes, whether for diabetes in general or for subpopulations of patients with diabetes (eg, patients with type 2 diabetes), nor has a review consolidated the revisions made to existing PROMs for diabetes.

Moreover, existing systematic reviews have focused on the psychometric properties of only certain categories of diabetes PROMs (eg, PROMs evaluating only health-related quality of life measures [8], PROMs for diabetes self-care [9], or PROMs in patients with diabetes associated with foot and ankle pathologies [10]) or the use of PROMs/association of PROMs with diabetes and its complications [11], even though such PROMs may be validated and applicable to a wider population of diabetes patients; for example, the Mexican version of the Diabetes Foot-Care Behavior Scale was validated in a population of patients with type 2 diabetes and not limited to patients with foot and ankle pathologies [12].

Therefore, we aimed to conduct a systematic literature review to identify studies investigating the measurement properties of PROMs validated in the population of patients with diabetes and evaluate the methodological quality and level of evidence relating to these measurement properties of PROMs. In addition, we aimed to categorize the PROMs by the type of outcome measure. This paper contains the psychometric results of the PROMs identified for patients with type 2 diabetes, and it is part of a series of papers to be published that will contain the results of the PROMs validated for (1) patients with type 1 diabetes; (2) patients with either type 1 or type 2 diabetes; and (3) patients with diabetes associated with complications such as peripheral neuropathy, retinopathy, or foot and ankle pathologies.

Methods

Review

This systematic review was guided by the PRISMA (Preferred Reporting Items for Systematic Review and Meta-Analyses) statement [13]. The measurement properties of each PROM were evaluated using the COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) Risk of Bias checklist [14]. The COSMIN evaluates PROM development and the following 9 measurement properties: content validity, structural validity, internal consistency, cross-cultural validity/measurement invariance, test-retest reliability, measurement error, criterion validity, hypotheses testing for construct validity, and responsiveness [14,15]. The results were used to determine the overall evidence of each PROM [16]. This systematic review has been submitted for registration on Prospero and registered on the Open Science Framework [17].

Search Strategy

The PubMed, Embase, and PsychINFO (Ovid) databases were searched for any articles published on or before March 31, 2020. A search strategy (Tables S1-S3, Multimedia Appendix 1) of three components was used as follows [18]: disease terms (diabetes and associated terms), construct of interest (PROMs and associated variations of this term), and measurement properties (as defined under the COSMIN criteria). Where available, the sensitivity of the searches was enhanced using search filters developed by Terwee et al [19] and the PROM Group, University of Oxford [20]. The search records were downloaded into Endnote X9 (Clarivate Analytics), and any duplicates were removed.

Article Selection

Two reviewers (PWJL and DHFL) independently screened all titles and abstracts, and a third reviewer (YHK) was consulted to make a final decision when any disagreement arose between the two reviewers as to the relevance of the articles based on the inclusion and exclusion criteria. For articles that were potentially relevant, the full-text articles were independently reviewed by the same two reviewers for inclusion and exclusion.

We included full-text original publications in English that validated PROMs for patients with diabetes mellitus and evaluated the PROMs for at least one of the nine measurement properties listed in the COSMIN guidelines. The COSMIN guidelines evaluate PROM development using the following nine measurement properties: content validity [21], structural validity, internal consistency, cross-cultural validity/measurement invariance, test-retest reliability, measurement error, criterion validity, hypotheses testing for construct validity, and responsiveness. Their definitions are presented by Mokkink et al [22].

We excluded conference abstracts and studies that focused on measurement development or that included PROMs completed by proxy or by patients ≤18 years of age. These exclusions were not used to construct the search strategy to avoid the omission of relevant studies. If only part of the study population consisted of PROMs directly reported by patients >18 years of age with diabetes, the articles were included if the results were reported separately for this group of patients. The type of study (eg, randomized controlled trial, cross-sectional study, cohort study, and registry-based study) was not part of our exclusion criteria to assess the measurement properties to ensure that this systematic literature review would be able to provide a comprehensive overview of the measurement properties of all types of PROMs.

Data Extraction

Two reviewers (PWJL and DHFL) extracted the following data (where available) from the articles:

1. General characteristics of the study populations: sample size, age, gender, and country where the study was conducted
2. Disease characteristics of the study population: disease studied and duration of illness
3. Characteristics of the PROMs: language version used, domains assessed, number of domains and items, and response scale

Assessment of Methodological Quality

Two reviewers (PWJL and DHFL) independently evaluated all relevant articles for methodological quality using the COSMIN Risk of Bias checklist [14], and a third reviewer (YHK) resolved any disagreement. Each measurement property was assessed based on a 4-point scale: inadequate, doubtful, adequate, or very good [14,15]. The item with the worst rating under each measurement property would determine the overall rating for the specific measurement property [23]. The assessed PROMs were then categorized according to their outcome measures.

Assessment of Quality of Measurement Properties

The quality of measurement properties of each PROM was assessed using the quality criteria described by Terwee et al [16]. First, the measurement properties to be evaluated were identified. Next, according to the results from the study of each measurement property, a “positive (+),” “indeterminate (?),” or “negative (-)” rating was assigned [16].

Evidence Synthesis

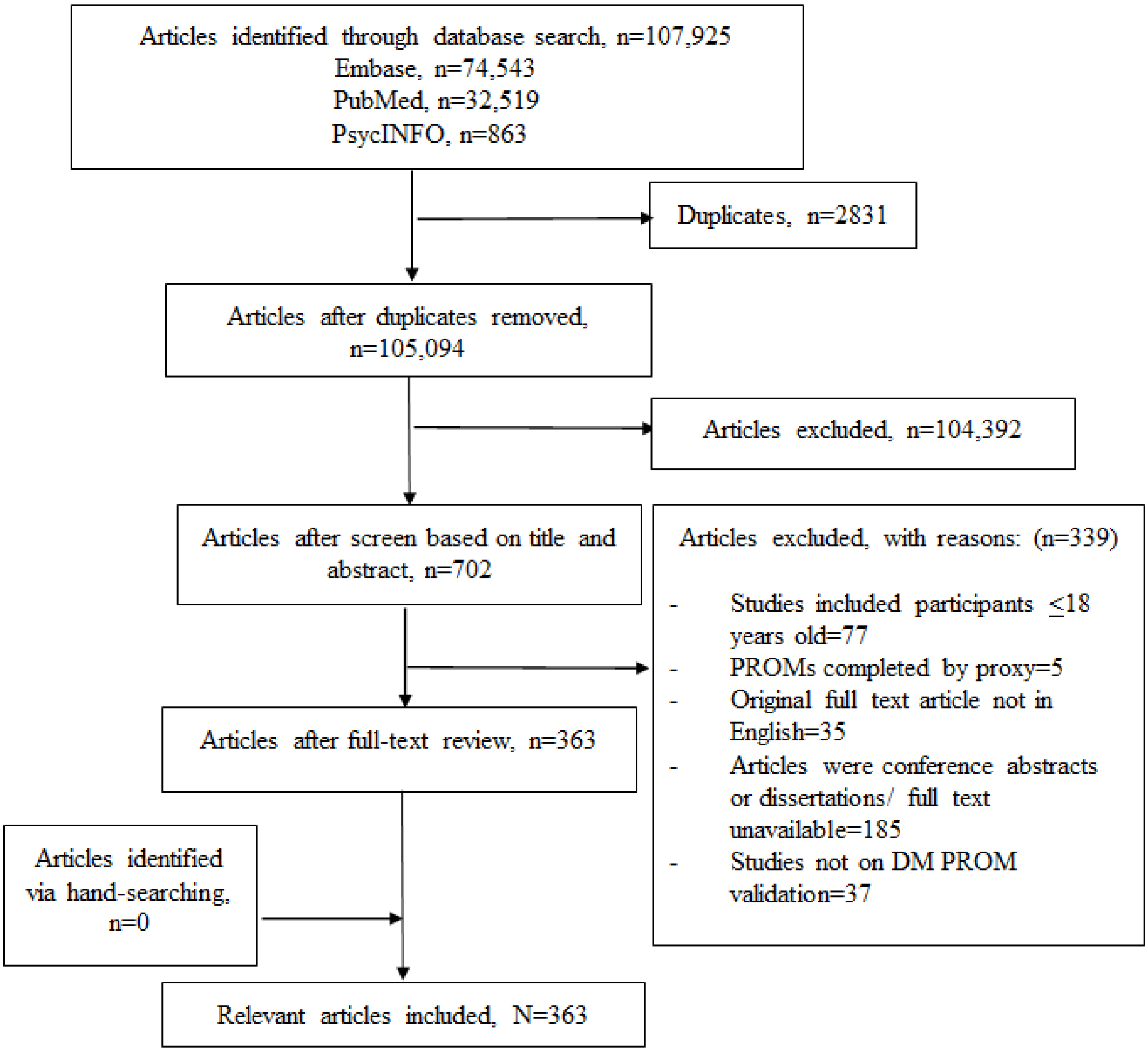
For each PROM, an evidence synthesis across all studies was conducted. First, we determined whether each measurement property for a PROM had overall “sufficient (+),” “insufficient (-),” “inconsistent (±),” or “indeterminate (?)” evidence. Second, we graded the quality of evidence for each measurement property of the PROM as high, moderate, low, or very low based on the guidelines from the modified Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach for systematic reviews of clinical trials [15,24].

Results

Search Results

A total of 107,925 articles were obtained from the database search (Figure 1), of which 2831 duplicates were excluded. A review of the titles and abstracts excluded 104,392 articles. Then, after a full-text review, 339 articles were excluded for the reasons provided in Figure 1, resulting in 363 relevant articles.

Figure 1. Flow chart of the systematic literature review. DM: diabetes mellitus; PROM: patient-reported outcome measure.


Articles identified through database search, n=107,925
Embase, n=74,543
PubMed, n=32,519
PsycINFO, n=863
Duplicates, n=2831
Articles after duplicates removed, n=105,094
Articles excluded, n=104,392
Articles after screen based on title and abstract, n=702
Articles excluded, with reasons: (n=339)
- Studies included participants ≤18 years old=77
- PROMs completed by proxy=5
- Original full text article not in English=35
- Articles were conference abstracts or dissertations/ full text unavailable=185
- Studies not on DM PROM validation=37
Articles after full-text review, n=363
Articles identified via hand-searching, n=0
Relevant articles included, N=363

Out of the 363 relevant articles, 209 articles reporting on 248 studies validated PROMs for patients with type 2 diabetes, which are the focus of the subsequent analysis in this paper. A breakdown of the 363 relevant articles identified by patient population is provided in Table 1. The results of studies assessing other diabetes subpopulations will be published elsewhere.

Out of the remaining 209 articles reporting on studies that validated PROMs for patients with type 2 diabetes, 238 unique PROMs in 35 languages from 53 countries were identified (Table 2).

Table 1. Breakdown of relevant articles by patient population (N=363).

Patient population	Articles, n (%)
Type 2 diabetes	209 (57.6)
Type 1 diabetes	16 (4.4)
Type 1 or Type 2 diabetes[a]	119 (32.9)
Diabetes with complications	19 (5.2)

[a] Includes articles that did not differentiate between types of diabetes. Attempts were made to contact the authors for clarification.

Table 2. Characteristics of the included articles.

General characteristics	Value
Unique PROMs[a] identified, n	238
Unique countries identified, n	53[b]
Unique languages identified, n	35[c]
Sample size,[d] n (%)	
<30	6 (2.42)
30-49	6 (2.42)
50-99	24 (9.68)
>100	203 (81.85)
Mean age[d] (years), n (%)	
30-39	3 (1.21)
40-49	10 (4.03)
50-59	108 (43.55)
60-69	80 (32.26)
≥70	11 (4.44)
Proportion of males,[e] n (%)	
<0.5	114 (45.97)
0.5 < x <0.6	83 (33.47)
0.6 < x < 0.7	25 (10.08)
0.7 < x < 0.8	3 (1.21)
0.8 < x < 0.9	2 (0.81)
Disease characteristics: disease duration (years),[e] n (%)	
0 < mean disease duration < 10	73 (29.44)
10 < mean disease duration < 20	67 (27.02)
20 < mean disease duration < 30	1 (0.40)

[a] PROMs: patient-reported outcome measures.

[b] Some countries were not reported.

[c] Some languages were not reported.

[d] Inclusive of multiple studies reported on the same sample.

[e] Some values were reported as median and range or were not reported.

Characteristics of the PROMs

The characteristics of the identified PROMs are presented in Table S4 in Multimedia Appendix 1. A majority of the PROMs studied were in English (27.82%), and all PROMs identified were self-administered questionnaires.

Categories of PROMs

The 238 unique PROMs identified are categorized in Table 3. Based on the intended outcome measurements as described by the authors of the respective validation studies, the PROMs can be broadly categorized into three groups: first, general impact on quality of life questionnaires (eg, the World Health Organization Quality of Life questionnaire [WHOQOL-100] and the EuroQol 5-Dimension [EQ-5D]) (24/238,10.1%); second, questionnaires measuring diabetes-specific impacts on quality of life (eg, the 19-item Audit of Diabetes-Dependent Quality of Life [ADDQoL-19]) (42/238,17.6%), and third, questionnaires measuring specific aspects of dealing with diabetes (eg, PAID, which measures psychological impact, or the Diabetes Treatment Satisfaction Questionnaire [DTSQ], which measures satisfaction with diabetes treatment) (172/238,72.3%). The majority of the PROMs fell into the third category, and they could be further classified into as follows: general psychosocial impact (eg, social/psychological/emotional well-being of patients with diabetes), diabetes-related depression, diabetes-related distress, self-efficacy (eg, patients' belief in their capability to organize and execute the course of action required to deal with their disease), self-management (eg, the range of activities patients must engage in on a regular basis to manage their diabetes), impact of empowerment tools (eg, the level of empowerment developed by patients as a result of educational interventions), health-promoting lifestyle behaviors (eg, in terms of physical activity or stress management), health beliefs (eg, perceived benefits of treatment), knowledge/competence, treatment experience (eg, the level of satisfaction with treatment), treatment compliance, symptoms (eg, hypoglycemia and patients' experiences/perceptions on this), nutrition and physical activity (eg, patients' perceptions on diet and exercise), sleep, support (eg, patients' perspectives on the availability of support for diabetes), attitude/coping with diabetes, obstacles and problem-solving, health perception (eg, patients' perspectives on their illness/diabetes-related health satisfaction).

Table 3. Patient-reported outcome measures (PROMs) organized by category.

Category	Description	PROMs[a]
General impact on quality of life	• Generic PROMs that are evaluated within the population of patients with type 2 diabetes. These PROMs assess the impact of chronic illness in terms of impact on quality of life. • Examples of domains assessed include physical health, psychological state, social relationships and environment, mobility, self-care, usual activities, pain/discomfort, and anxiety/depression.	WHOQOL-100[b] [25]; WHOQOL-BREF[c] [26]; WHOQOL-BREF (Malayalam version) [27]; WHOQOL-BREF 26 (Persian version) [28]; WHOQOL-BREF (Amharic version) [29]; RAND-12[d] [30]; HSQ 2.0[e] (Spanish version) [31]; HUI2[f] [30]; HUI3[g] [30,32]; EQ-5D[h] [33], [34]; EQ-5D-5L[i] [35-37]; EQ-5D-3L[j] [38]; EQ-5D-3L (Finnish version) [38]; EQ-5D-3L (German version) [38]; EQ-5D-3L (Greek version) [38]; EQ-5D-3L (Dutch version) [38]; EQ-5D-3L (Spanish version) [38]; EQ-5D-5L (Brunei-Malay version) [39]; PACIC[k] [40,41]; Short-version PACIC [42]; Modified-PACIC [43]; PACIC (Malay version) [44]; SF-36[l] [45]; SF-12v2[m] [46]
Diabetes-specific impact on quality of life	• PROMs that assess the impact of diabetes on quality of life, such as impact on physical function, psychological well-being and social well-being. • Examples of domains assessed include physical function, symptoms, psychological well-being, self-care management, social well-being, global judgments of health, and satisfaction with care and flexibility of treatment.	PRO-DM-Thai[n] [47]; DQOL[o] (Chinese version) [48,49]; DQOL (Iranian version) [50]; DQOL (Turkish version) [51]; DQOL (Malay version) [52]; IRDQOL[p] [28]; revised version of DQOL [53]; AsianDQOL[q] [54]; AsianDQOL (Malay version) [54]; AsianDQOL (Chinese-mandarin version) [54]; DQL-BCI[r] (Polish version) [55]; [s]DQOL-B [56]; QOLID[t] [57]; J-DQOL[u] [58]; QOL[v] questionnaire [59]; DMQoL[w] (Persian version) [60]; MENQOL[x] [61]; Diabetes-39 (Arabic version) [62]; Diabetes-39 (Brazillian version) [63]; ADDQoL-19[y] [56], [64-66]; ADDQoL-19 (Chinese version) [67]; ADDQoL-19 (Malay version) [64]; CN-ADDQoL[z] [68]; ADDQoL (Spanish version) [69]; ADDQoL (Turkish version) [70]; Malay ADDQoL [71]; Elasy et al [72]; DHP-1[aa] [73]; DHP-3D[ab] [74]; DHP-5D[ac] [74]; DCP[ad] (Chinese version) [75]; DIMS[ae] (Chinese version) [76]
General psychosocial impact	• PROMs that assess the social/psychological/emotional well-being of patients with diabetes. • Examples of domains assessed include anxiety, depressed mood, positive well-being, self-control, general health, and validity.	MDQ[af] [77]; MDQ (Hindi version) [78]; PGWB[ag] [33]; WBQ[ah] [26,79]; W-BQ28[ai] [80]; WHO-5[aj] [81]; WHO-5 (Polish version) [82]
Diabetes-related depression	• PROMs that screen for depression in patients with diabetes/monitor the presence of depressive symptoms in patients with diabetes. • Examples of domains assessed include depressed affect, somatic symptoms, positive affect, and interpersonal problems.	CES-Depression[ak] [83-86]; Depression in Diabetes Self-Rating Scale [87]; SCAD[al] [84]; HADS[am] [84]; DMI[an] [84]; EDS[ao] [88]; DCS[ap] [89]; CUDOS-Chinese[aq] [90]; PHQ-9[ar] [91], [92]; PHQ-9 (Chichewa version) [93]; PHQ-9 (Romanian version) [94]
Diabetes-related distress	• PROMs that screen for diabetes-related emotional distress. • Examples of domains assessed include emotional burden, physician-related distress, regimen-related distress, and interpersonal distress.	CDDS-17[as] [95]; DDS Bahasa Indonesia[at] [96]; PAID[au] [97]; MY-PAID-20[av] [98]; B-PAID[aw] [99]; PAID-K[ax] [100]; K-PAID[ay] [101]; K-PAID-5[az] [101]; Turkish PAID [102]; PAID (Greek version) [103]; SG-PAID-C[ba] [104]; PAID (Spanish version) [105]; IR-PAID-20[bb] [106]
Self-efficacy	• PROMs that assess the level of self-efficacy (ie, people's belief in their capability to organize and execute the courses of action required to deal with prospective situations [107]) of patients with type 2 diabetes, whether in general or in dealing with specific aspects of diabetes (eg, in taking medication [108]). • Examples of domains assessed include performing activities which are essential for the treatment of diabetes, self-observation, and self-regulating activities.	SE-Type 2[bc] [107]; DMSES[bd] [109]; K-DMSES[be] [110]; GR-DMSES[bf] [111]; DMSES (Brazilian version) [112]; IT-DMSES[bg] [113]; DSEQ[bh] (Thai version) [114]; CDMSS-11[bi] [115]; DSCAS[bj] [116]; DSES[bk] [116]; K-DSES[bl] [117]; Situational Self-Efficacy Scales (Spanish version) [31]; ESS[bm] [118]; Self-Efficacy for Exercise 1 (Spanish version) [31]; Self-Efficacy for Exercise 2 (Spanish version) [31]; PTES[bn] [108]

Category	Description	PROMs[a]
Self-management	- PROMs that assess the level of diabetes self-management (ie, range of activities in which individuals must engage on a regular basis to manage their diabetes [119]). - Examples of domains assessed include general diet, specific diet, exercise, medication taking, blood-glucose testing, foot care, and cigarette smoking.	SDSCA[bo] [120]; SDSCA (Turkish version) [121]; SDSCA-G[bp] [122]; SDSCA (Moroccan version) [123]; SDSCA-Ar[bq] [124]; SDSCA-K[br] [125]; INAAP-DM2[bs] [126]; SCI-R[bt] [127]; DSSCI[bu] [128]; SUGAAR[bv] [129]; D-SMART[bw] [119]; ES-SMBPA-2D[bx] [130]; DSMS[by] [116]; DSMQ[bz] (Thai version) [131]; DSMQ (Urdu version) [132]; V-DSMI[ca] [133]; DSMI-20[cb] [134]; DSMB-O[cc] [135]; SMP-T2D[cd] [136]; PAM13[ce] [137]; Chernyak et al [138]; CIRS[cf] (Thai version) [139]
Impact of empowerment tools	- PROMs that assess the level of empowerment (ie, patients' natural capacity and ability to become responsible for their own lives) that is discovered and developed [140] as a result of educational interventions. - Examples of domains assessed include managing the psychosocial aspects of diabetes, assessing dissatisfaction, and readiness to change.	IR-DES-28[cg] [140]; Hara et al [141]; DES-M[ch] [142]; DES-SF[ci] (Brazilian Portuguese version) [143]; DES-SF (Portuguese version) [144]
Health-promoting lifestyle behaviors	- PROMs that assess health-promoting lifestyle behaviours of patients with diabetes. - Examples of domains assessed include physical activity, risk reduction, stress management, health responsibility, enjoyment of life, and healthy diet.	T2DHPS[cj] (Persian version) [145]; T2DHPS (Turkish version) [146]; DHPSC[ck] (Chinese version) [147]; PDQ-11[cl] [148]; C-PDQ[cm] [149]
Health beliefs	- PROMs that assess diabetes-specific health beliefs of patients. - Examples of domains assessed include perceived benefits of and barriers to treatment and perceived severity of and vulnerability to complications.	Health Belief Measures [150]; Given Health Belief Instrument (Spanish version) [151]; Health Belief Model Scale (Turkish version) [152]; Diabetes Health Belief Measure [153]
Knowledge/ competence	- PROMs that assess the level of diabetes knowledge, whether in general or for specific areas of knowledge such as nutrition knowledge. - Examples of domains assessed include symptoms (eg, frequent hunger), causes and risk factors (eg, lack of physical activity), complications (eg, kidney failure), and management (eg, reduced consumption of rice).	Diabetes Questionnaire [154]; Diabetes Questionnaire (Spanish version) [154]; Diabetes Knowledge Questionnaire (Spanish version) [31]; DKQ-24[cn] [153]; DMKT[co] [155]; PCSD-P[cp] [156]; Miller et al [157]; Miller and Edwards [158]; PDDC[cq] [159]; DRNK[cr] [160]; FCCHL[cs] (Norwegian version) [161]; KHLS-DM[ct] [162]; HLS-K[cu] [163]; HLS/SNS[cv] [164]; Ashok et al 1 [165]; Ashok et al 2 [166]; HLS-EU-Q47[cw] [167]
Treatment experience	- PROMs that assess the treatment experience in general [168] or specifically the level of satisfaction with treatment [79] or treatment burden [169] or treatment with specific modalities of treatment (eg, with pharmacotherapy [169] or insulin therapy [170]). - Examples of domains assessed include efficacy, treatment burden and symptoms (side effects), diabetes worries, perceptions of insulin therapy, treatment satisfaction, and inhaler performance.	DTSQ[cx] [79]; DTSQ (Greek version) [171]; DiabMedSat[cy] [172]; DTBQ[cz] [169]; ITEQ[da] [168]; IITQ[db] [170]; ITAS[dc] [173]; C-ITAS-HK[dd] [174]; BITQ[de] (Turkish version) [175]; Ch-ASIQ[df] [176]; MIAS[dg] [177]; IMDSES[dh] (Brazilian version) [178]; ITSQ[di] [179]; OHA-Q[dj] [180]; DMSRQ[dk] [181]
Treatment compliance	- PROMs that measure the level of compliance to treatment/adherence to medication/patients' beliefs regarding treatment. - Examples of domains assessed include emotional difficulties in compliance, physical difficulties in compliance, changing difficulties of habits in compliance, acceptance difficulties in compliance, awareness difficulties in compliance, diet difficulties in compliance, and denial difficulties in compliance.	Demirtas et al [182]; MMAS[dl] (Thai version) [183]; modified 4-item Morisky–Green–Levine Medication Adherence Scale [184]; MMAS-8[dl] (Korean version) [185]; MMAS-8 (Chinese version) [186]; MMAS-8 (French version) [187], [188]; MGLS[dm] (Indonesian version) [189]; Medical Prescription Knowledge questionnaire [190]; Attitude Scale [190]; BMQ-f[dn] [191]; MALMAS[do] [192]; MAT OADs[dp] [193]; MAT Insulin[dq] [193]; ARMS-K[dr] [194]; Diabetes Medication System Rating Questionnaire Short-Form [195]; SR-4[ds] (French version) [187]; Zongo et al 1 [187]; Zongo et al 2 [187]
Symptoms		

Category	Description	PROMs[a]
	• PROMs that assess patients' experiences/perceptions of specific symptoms associated with diabetes (eg, hypoglycemia [196], fatigue [197]). • Examples of domains assessed include symptom concern, compensatory behavior, worry, general fatigue, and physical fatigue.	HPQ[dt] (Cyprus version) [196]; HPQ [196]; CHI[du] (Filipino version) [198]; FH-15[dv] (Chinese version) [199]; K-DSC-R[dw] [200]; DSC-R[dx] [201]; Naegeli et al [202]; FACIT[dy]-Fatigue Scale [197]
Nutrition and physical activity	• PROMs that assess patients' perspectives (eg, barriers/confidence level/knowledge) in relation to diet/nutrition and exercise/physical activity. • Examples of domains assessed include satisfaction with diet, burden of diet therapy, perceived merits of diet therapy, general perception of diet, restriction of social functions, vitality, and mental health.	Barriers to Fat Reduction Scale[dz] (Spanish version) [31]; Barriers to Exercise Checklist (Spanish version) [31]; Food Habits Questionnaire (Spanish version) [31]; DDRQOL[dz] [203]; DDRQOL-R[ea] [204]; Sato et al [204]; IW-SP[eb] [205]; Motiva.Diaf-DM2 questionnaire [206]; HAPA-based PA inventory[ec] [207]
Sleep	• PROMs that assess patients' sleep symptoms in general or specific sleep-related issues, such as obstructive sleep apnea symptoms [208].	STOP-Bang questionnaire [208]; PROMIS[ed]–Sleep Disturbance instrument [209]; PROMIS–Sleep Related Impairment instrument [209]
Support	• PROMs that assess patients' perspectives on availability of resources/support for the management of diabetes. • Examples of domains assessed include individualized assessment, collaborative goal setting, enhancing skills, ongoing follow-up and support, and community resources.	RSSM-Farsi[ee] [210]; DFBC[ef] (Japanese version) [211]; FSS-AA T2DM[eg] [212]; HCCQ-P[eh] [213]; The Diabetes Family Support and Conflict Scale (Turkish version) [214]
Attitude/coping with diabetes	• PROMs that assess perception toward disease, such as self-stigma (patients' own negative attitude toward themselves [215]), relationship consciousness [216] or awareness of the psychological burden of disease [217]. • Examples of domains assessed include cognitive, affective, behavioral, psychological impact of diabetes, sense of self-control, and efforts for symptom management.	SSS-J[ei] [215]; DSAS-2[ej] [218]; Relationship Consciousness of Japanese Patients with Type 2 Diabetes Mellitus Scale [216]; ADS[ek] (Japanese version) [217]; ADS (Korean version) [219]; DAAS[el] [220]; IR-DAS-3[em] [221]; GCQ[en] [222]; DIAB-Q[eo] [223]; S-BRCS[ep] [224]
Obstacles and problem-solving	• PROMs that assess patients' perspectives on obstacles to self-management/approach to manage problems in diabetes self-management/ desire to participate in medical decision-making. • Examples of domains assessed include desire for discussion and desire for information, medication, self-monitoring, knowledge and beliefs, diagnosis, relationships with health care professionals, lifestyle changes, coping, and advice and support.	DPMD[eq] [225]; DOQ[er] [226]; DOQ (Dutch version) [227]; DOQ-30[es] [228]; DPSS[et] [229]
Health perception	• PROMs that assess patients' general perceptions on their illness/diabetes-related health satisfaction and knowledge of the disease or, specifically, the perception of fatalism (events are fixed such that humans are powerless to change them) [230]. • Examples of domains assessed include timeline-acute/chronic, consequences, personal control, treatment control, illness coherence, emotional representation, and cause component.	IPQ-R[eu] [231]; CHES-Q[ev] [232]; MBIPQ[ew] [233]; DFS[ex] [230]

---

[a] PROMs: patient-reported outcome measures.

[b] WHOQOL-100: World Health Organization Quality of Life questionnaire.

[c] WHOQOL-BREF: abbreviated World Health Organization Quality of Life questionnaire.

[d] RAND-12: Veterans RAND 12-Item Health Survey.

[e] HSQ 2.0: Health Status Questionnaire 2.0.

[f] HUI2: Health Utilities Index Mark 2.

[g] HUI3: Health Utilities Index Mark 3.

[h] EQ-5D: EuroQol 5-Dimension.

[i] EQ-5D-5L: EuroQol 5-Dimension with 5-level scale.

[j] EQ-5D-3L: EuroQol 5-Dimension with 3-level scale.

[k] PACIC: Patient Assessment of Chronic Illness Care.

[l] SF-36: 36-Item Short Form Survey.

[m] SF-12v2: Short Form-12 Health Survey version 2.

[n] PRO-DM-Thai: instrument for patient-reported outcomes in Thai patients with type 2 diabetes mellitus.

[o] DQOL: Diabetes Quality-of-Life Measure.

[p] IRDQOL: Iranian Diabetes Quality of Life.

[q] AsianDQOL: Asian Diabetes Quality of Life.

[r] DQL-BCI: Diabetes Quality of Life–Brief Clinical Inventory.

[s] DQOL-B: Diabetes Quality of Life Brief Clinical Inventory.

[t] QOLID: Quality of Life Instrument for Indian Diabetes Patients.

[u] J-DQOL: Japanese version of the Diabetes Quality-Of-Life Measure.

[v] QOL: quality of life.

[w] DMQoL: Diabetes-Mellitus Specific Quality of Life.

[x] MENQOL: Menopause-specific Quality of Life.

[y] ADDQoL-19: 19-item Audit of Diabetes-Dependent Quality of Life.

[z] CN-ADDQoL: Adaptation of the ADDQoL questionnaire to people with diabetes in China.

[aa] DHP-1: Diabetes Health Profile.

[ab] DHP-3D: Diabetes Health Profile–3 Dimension.

[ac] DHP-5D: Diabetes Health Profile–5 Dimension.

[ad] DCP: Diabetes Care Profile.

[ae] DIMS: diabetes impact measurement scales.

[af] MDQ: Multidimensional Diabetes Questionnaire.

[ag] PGWB: Psychological General Well-Being Questionnaire.

[ah] WBQ: Well-being Questionnaire.

[ai] W-BQ28: 28-item Well-Being Questionnaire.

[aj] WHO-5: 5-item World Health Organization well-being index.

[ak] CES-Depression: Center for Epidemiological Studies Depression scale.

[al] SCAD: Silverstone Concise Assessment for Depression.

[am] HADS: Hospital Anxiety and Depression Scale.

[an] DMI: Depression in the Medically Ill Questionnaire.

[ao] EDS: Edinburgh Depression Scale.

[ap] DCS: Depressive Cognition Scale.

[aq] CUDOS-Chinese: Mandarin Chinese Version of the Clinically Useful Depression Outcome Scale.

[ar] PHQ-9: Patient Health Questionnaire-9.

[as] CDDS-17: Chinese version of the Diabetes Distress Scale.

[at] DDS Bahasa Indonesia: Indonesian Diabetes Distress Scale.

[au] PAID: Problem Areas in Diabetes scale.

[av] MY-PAID-20: Malaysian version of the Problem Areas in Diabetes scale.

[aw] B-PAID: Brazilian version of the Problem Areas in Diabetes scale.

[ax] PAID-K: Korean version of the Problem Areas in Diabetes scale.

[ay] K-PAID: Korean translation of the Problem Areas in Diabetes scale.

[az] K-PAID-5: Korean translation of the short form Problem Areas in Diabetes scale.

[ba] SG-PAID-c Chinese version of the Problem Areas in Diabetes Scale.

[bb] IR-PAID-20: Iranian version of the Problem Areas in Diabetes Scale.

[bc] SE-Type 2: self-efficacy scale for patients with type 2 diabetes mellitus.

[bd] DMSES: diabetes management self-efficacy scale.

[be] K-DMSES: Korean version of the diabetes management self-efficacy scale.

[bf] GR-DMSES: Greek version of the diabetes management self-efficacy scale.

[bg] IT-DMSES: Italian version of the diabetes management self-efficacy scale.

[bh] DSEQ: Self-Efficacy for Diabetes Scale.

[bi] CDMSS-11: Chinese version of the Diabetes Medication Self-efficacy Scale.

[bj] DSCAS: Diabetes Self-Care Agency Scale.

[bk] DSES: Diabetes Self-efficacy Scale.

[bl] K-DSES: Korean version of the Diabetes Self-efficacy Scale.

[bm] ESS: Exercise Self-efficacy Scale.

[bn] PTES: Perceived Therapeutic Efficacy Scale.

[bo] SDSCA: Summary of diabetes self-care activities measure.

[bp] SDSCA-G: German version of the Summary of diabetes self-care activities measure.

[bq] SDSCA-Ar: Arabic version of the Summary of diabetes self-care activities measure.

[br] SDSCA-K: Korean version of the Summary of diabetes self-care activities measure.

[bs] INAAP-DM2: Self-care Assessment Instrument for patients with type 2 diabetes mellitus.

[bt] SCI-R: Self-Care Inventory-Revised.

[bu] DSSCI: Diabetes Symptom Self-Care Inventory.

[bv] SUGAAR: Self-Care Utility Geriatric African-American Rating.

[bw] D-SMART: Diabetes Self-management Assessment Report Tool.

[bx] ES-SMBPA-2D: evaluation scale for self-management behavior related to physical activity of type 2 diabetic patients.

[by] DSMS: Diabetes Self-Management Scale.

[bz] DSMQ: Diabetes Self-management Questionnaire.

[ca] V-DSMI: Vietnamese version of the Diabetes Self-Management Instrument.

[cb] DSMI-20: Diabetes Self-Management Instrument Short Form.

[cc] DSMB-O: Diabetes Self-Management Behavior for Older Koreans.

[cd] SMP-T2D: self-management profile for type 2 diabetes.

[ce] PAM13: Patient Activation Measure 13.

[cf] CIRS: Chronic Illness Resources Survey.

[cg] IR-DES-28: Iranian version of the Diabetes Empowerment Scale.

[ch] DES-M: diabetes empowerment scale.

[ci] DES-SF: Diabetes Empowerment Scale–Short Form.

[cj] T2DHPS: Type 2 Diabetes and Health Promotion Scale.

[ck] DHPSC: diabetes health promotion self-care scale.

[cl] PDQ-11: Personal Diabetes Questionnaire.

[cm] C-PDQ: Chinese version of the Personal Diabetes Questionnaire.

[cn] DKQ-24: Diabetes Knowledge Questionnaire-24.

[co] DMKT: Diabetes Mellitus Knowledge Test.

[cp] PCSD-P: Persian Version of the Perceived Competence Scale for Diabetes.

[cq] PDDC: measure of perceived diabetes and dietary competence.

[cr] DRNK: diabetes-related nutrition knowledge questionnaire.

[cs] FCCHL: Functional, Communicative, and Critical Health Literacy Scale.

[ct] KHLS-DM: Korean Health Literacy Scale for Diabetes Mellitus.

[cu] HLS-K: Health Literacy Scale.

[cv] HLS/SNS: Health Literacy Scale/Subjective Numeracy Scale.

[cw] HLS-EU-Q47: European Health Literacy Survey Questionnaire.

[cx] DTSQ: Diabetes Treatment Satisfaction Questionnaire.

[cy] DiabMedSat: Diabetes Medication Satisfaction measure.

[cz] DTBQ: Diabetic Treatment Burden Questionnaire.

[da] ITEQ: insulin treatment experience questionnaire.

[db] IITQ: inhaled insulin treatment questionnaire.

[dc] ITAS: Insulin Treatment Appraisal Scale.

[dd] C-ITAS-HK: Hong Kong version of the Chinese Insulin Treatment Appraisal Scale.

[de] BITQ: Barriers to Insulin Treatment Questionnaire.

[df] Ch-ASIQ: Chinese Attitudes to Starting Insulin Questionnaire.

[dg] MIAS: Morisky Medication Adherence Scale adapted to specify insulin adherence.

[dh] IMDSES: Insulin Management Diabetes Self-Efficacy Scale.

[di] ITSQ: Insulin Treatment Satisfaction Questionnaire.

[dj] OHA-Q: Oral Hypoglycemic Agent Questionnaire.

[dk] DMSRQ: Diabetes Medication System Rating Questionnaire.

[dl] MMAS-8: 8-item Morisky Medication Adherence Scale.

[dm] MGLS: 4-item Morisky-Green-Levine Adherence Scale.

[dn] BMQ-f: French version of the Beliefs about Medicines Questionnaire.

[do] MALMAS: Malaysian Medication Adherence Scale.

[dp] MAT OADs: Measurement of Adherence to Drug Therapy in Diabetes Mellitus–Oral Antidiabetics.

[dq] MAT Insulin: Measurement of Adherence to Drug Therapy in Diabetes Mellitus–Insulin Therapy.

[dr] ARMS-K: Korean version of the Adherence to Refills and Medications Scale.

[ds] SR-4: self-report with 4 items.

[dt] HPQ: Hypoglycemia Perspectives Questionnaire.

[du] CHI: Clarke Hypoglycemia Index.

[dv] FH-15: Chinese version of the new Fear of Hypoglycemia scale.

[dw] K-DSC-R: Korean version of the Diabetes Symptom Checklist-Revised.

[dx] DSC-R: Diabetes Symptom Checklist-Revised.

[dy] FACIT: Functional Assessment of Chronic Illness Therapy.

[dz] DDRQOL: Diabetes Diet-Related Quality-of-Life scale.

[ea] DDRQOL-R: revised and short form versions of the Diabetes Diet-Related Quality of Life scale.

[eb] IW-SP: Impact of Weight on Self-Perceptions Questionnaire.

[ec] HAPA-based PA inventory: health action process approach (HAPA)–based physical activity inventory.

[ed] PROMIS: Patient-Reported Outcomes Measurement Information System.

[ee] RSSM-Farsi: Iranian version of Resources and Support for Chronic Illness Self-management scale.

[ef] DFBC: Diabetes Family Behavior Checklist.

[eg] FSS-AA T2DM: Family Support Scale Adapted for African American Women with Type 2 Diabetes Mellitus.

[eh] HCCQ-P: Persian Health Care Climate Questionnaire.

[ei] SSS-J: Japanese version of the Self-Stigma Scale.

[ej] DSAS-2: Type 2 Diabetes Stigma Assessment Scale.

[ek] ADS: Appraisal of Diabetes Scale.

[el] DAAS: Diabetes Adjustment Assessment Scale.

[em] IR-DAS-3: Iranian Diabetes Attitude Scale.

[en] GCQ: General Coping Questionnaire.

[eo] DIAB-Q: 17-item Diabetes Intention, Attitude, and Behavior Questionnaire.

[ep] S-BRCS: Spanish Brief Religious Coping Scale.

[eq] DPMD: diabetes-specific measure of patient desire to participate in medical decision making.

[er] DOQ: Diabetes Obstacles Questionnaire.

[es] DOQ-30: short version of the Diabetes Obstacles Questionnaire.

[et] DPSS: Diabetes Problem-Solving Scale.

[eu] IPQ-R: Revised Illness Perception Questionnaire.

[ev] CHES-Q: 14-item Current Health Satisfaction Questionnaire.

[ew] MBIPQ: Malay version of the Brief Illness Perception Questionnaire.

[ex] DFS: 12-item Diabetes Fatalism Scale.

Assessment of Methodological Quality and Quality of Measurement Properties

The results from the assessment of methodological quality and quality of measurement properties of the PROMs are presented in Table S5 of Multimedia Appendix 1. In terms of validity, hypothesis testing for construct validity, structural validity, and content validity were measured for 46.8% (116/248), 49.2% (122/248), and 29.0% (72/248) of the studies, respectively. In terms of reliability, internal consistency and reliability were assessed in 79.0% (196/248) and 41.9% (104/248) of the studies, respectively.

Evidence Synthesis

The results from the evidence synthesis of the PROMs are summarized in Table S6 in Multimedia Appendix 1. PROMs with at least a moderate level of evidence for ≥5 measurement properties include the Chinese version of the Personal Diabetes Questionnaire (C-PDQ) and the Insulin Treatment Appraisal Scale in Hong Kong primary care patients (C-ITAS-HK), of which the C-PDQ has a sufficient (+) rating for at least 4 measurement properties.

Recommendations

According to the COSMIN guidelines [15], PROMs that have evidence for sufficient content validity and at least low-quality evidence for sufficient internal consistency can be recommended for use, and the results obtained with these PROMs can be trusted. The 43 PROMs that meet these criteria are shaded in Table S6 in grey and presented in Table S7 (Multimedia Appendix 1). They are listed below according to the categorization we have proposed in Table 3:

1. General impact on quality of life: Health Status Questionnaire 2.0 (HSQ 2.0) (Spanish version), Patient Assessment of Chronic Illness Care (PACIC)
2. Diabetes-specific impact on quality of life: instrument for patient-reported outcomes in Thai patients with type 2

diabetes mellitus (PRO-DM-Thai), Diabetes Quality-of-Life Measure (DQOL), Asian DQOL

3. Diabetes-related depression: Mandarin Chinese Version of the Clinically Useful Depression Outcome Scale (CUDOS-Chinese), Patient Health Questionnaire-9 (PHQ-9).
4. Self-efficacy: diabetes management self-efficacy scale (DMSES), Situational Self-Efficacy Scales (Spanish version), Self-Efficacy for Exercise 1 (Spanish version), Self-Efficacy for Exercise 2 (Spanish version)
5. Self-management: Diabetes Self-management Questionnaire (DSMQ), Diabetes Self-Management Instrument Short Form (DSMI-20), Chronic Illness Resources Survey (CIRS) (Thai version)
6. Impact of empowerment tools: Diabetes Empowerment Scale–Short Form (DES-SF)
7. Lifestyle behaviors: Type 2 Diabetes and Health Promotion Scale (T2DHPS), diabetes health promotion self-care scale (DHPSC) (Chinese version), C-PDQ
8. Health beliefs: Health belief model (Turkish version)
9. Knowledge/competence: Diabetes Knowledge Questionnaire (Spanish version), Diabetes Mellitus Knowledge Test (DMKT), Persian Version of Perceived Competence Scale for Diabetes (PCSD-P), Miller et al [157], the diabetes-related nutrition knowledge questionnaire (DRNK), Korean Health Literacy Scale for Diabetes Mellitus (KHLS-DM).
10. Treatment experience: C-ITAS-HK, Chinese Attitudes to Starting Insulin Questionnaire (CH-ASIQ).
11. Treatment compliance: Medical Prescription Knowledge questionnaire, Attitude Scale, Measurement of Adherence to Drug Therapy in Diabetes Mellitus–Oral Antidiabetics (MAT OADS), Measurement of Adherence to Drug Therapy in Diabetes Mellitus–Insulin Therapy (MAT insulin).
12. Symptoms: new fear of hypoglycemia scale (FH-15) (Chinese version), Functional Assessment of Chronic Illness Therapy (FACIT)–Fatigue Scale.
13. Nutrition and physical activity: Barriers to Fat Reduction Scale (Spanish version), Barriers to Exercise Checklist (Spanish version), Food Habits Questionnaire (Spanish version), health action process approach (HAPA)–based PA inventory
14. Support: Persian Health Care Climate Questionnaire (HCCQ-P), the Diabetes Family Support and Conflict Scale (Turkish version)
15. Attitude/coping with diabetes: Relationship Consciousness of Japanese Patients with Type 2 Diabetes, Diabetes Adjustment Assessment Scale (DAAS)
16. Obstacles and problem-solving: diabetes-specific measure of patient desire to participate in medical decision making (DPMD), Diabetes Obstacles Questionnaire (DOQ)

Discussion

Principal Findings

To the best of our knowledge, this is the first systematic review to attempt to comprehensively summarize and categorize the PROMs for type 2 diabetes and to assess their overall level of evidence based on the COSMIN guidelines. Among the 248 included studies, we identified 238 unique PROMs for patients with type 2 diabetes, with 43 unique PROMs meeting the COSMIN guidelines for recommendation for use. While our study identified a wide range of unique PROMs, based on the number of studies for each PROM (in Table S6, Multimedia Appendix 1), most of the PROMs have been validated by very few studies that evaluated them in the Type 2 diabetes population, which may bias the assessment of methodological quality and quality of measurement properties.

Nevertheless, according to the COSMIN guidelines [15], PROMs that have been validated by at least one study showing sufficient content validity and at least low quality evidence for sufficient internal consistency can be recommended for use. Our review has included recommendations for various PROM categories that should be helpful for clinicians and academics.

As illustrated in Table 3, the 43 unique PROMs identified and recommended for use measure a wide range of clinically relevant domains, ranging from impact on quality of life to specific issues such as treatment experience/treatment compliance for clinicians and researchers to apply in clinical practice. Further validation studies can also be conducted on the remaining 189 PROMs that do not meet the COSMIN guidelines for recommendation.

Measurement error was assessed in only one study, as the other studies did not report standard error of measurement, smallest detectable change, or limits of agreement as required by the COSMIN. This may be addressed in future research. In addition, although PROM translations were performed for 122 out of the 248 studies (49.2%), none of these studies assessed measurement invariance or differential item functioning; therefore, cross-cultural validity was not evaluated for any of the PROMs in this study. Further studies on measurement error and cross-cultural validity of medication adherence PROM are warranted for these studies.

Strengths and Limitations

Our study has several strengths. We used three databases and sensitive search filters to capture as many potentially relevant articles as possible. The rigor of the study was established using the PRISMA statement and the COSMIN guidelines, which are well regarded as a consensus-based standard for evaluating the measurement properties of PROMs. The COSMIN Risk of Bias checklist employed in this study is an improvement from the original COSMIN checklist, with several improvements in the standards for evaluation [14,15]. As far as possible, we have also aimed to adopt the COSMIN guidelines for reporting the results of our evaluation. The PRISMA statement was used because it improves the transparency and clarity of the systematic review [234]. Moreover, in Table 3, we categorized all the PROMs based on their types of outcome measures. In addition to having highlighted 43 unique PROMs that are recommended for use under the COSMIN guidelines, the categorization we have proposed will provide readers with a range of options for selecting the most appropriate or robust PROM according to their required domain of assessment.

Our study has some limitations. One limitation related to this study is that the selection and evaluation of articles were subjective in nature and may have been prone to judgment bias. Further, given the scope of our study, there is inevitable potential for remaining inaccuracies in the data review/extraction process. Nevertheless, the requirement by COSMIN to have two independent reviewers and the need for a third reviewer to reach a consensus in the case of any discrepancy helps reduce the risk of judgement bias [15] and reduces the likelihood of any inaccuracies. Further, this study included only English full-text articles. Full-text articles were necessary, as they are peer-reviewed and recommended for inclusion by Terwee et al [235].

Further, in terms of the scope of our literature review, the PROMs identified may have been validated in other disease populations, or versions of the same PROMs may have been translated into other languages or culturally adapted to other populations; thus, they may have been validated in separate studies not captured by our literature review. For the validation studies of the PROMs that were identified, given the high number of studies retrieved, we were unable to hand-search all the references of the studies retrieved, and there may be additional relevant studies that were not included. Thus, such studies would be outside the scope of this review.

Finally, in comparison with previous systematic reviews on diabetes PROMs (ie, PROMs evaluating only health-related quality of life measures [8], PROMs for diabetes self-care [9], PROMs in patients with diabetes associated with foot and ankle pathologies [10], or the use of PROMs/association of PROMs with diabetes and its complications [11]), a direct comparison to the PROMs reviewed by these prior studies, which focused on diabetes complications, could not be made in this review, which focuses on reporting the results of the PROMs on type 2 diabetes (and not including complications of diabetes) identified from our review. The results of our systematic review for PROMs on the complications of diabetes will be reported separately.

Conclusion

This review has identified 238 unique PROMs for type 2 diabetes through a systematic review and evaluated their level of evidence, adjusted using results from an assessment of methodological quality. Based on the COSMIN guidelines for evidence synthesis, PROMs with at least a moderate level of evidence for ≥5 measurement properties include the C-PDQ, DSMI-20, and the C-ITAS-HK, of which the C-PDQ has sufficient (+) ratings for at least 4 measurement properties, and based on the COSMIN guidelines, 43 unique PROMs can be recommended for use.

Acknowledgments

We would like to thank Librarian Ms Wong Suei Nee at the National University of Singapore Medical Library for her advice on search strategies. This research is supported by the AM-ETHOS Duke-NUS Medical Student Fellowship Award and the Ministry of Health Singapore National Innovation Challenge Grant on Chronic Diseases Management (MOH/NIC/CDM1/2018).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary material.
[DOCX File , 291 KB - jmir_v23i8e25002_app1.docx]

References

1. IDF Diabetes Atlas, 9th edition. International Diabetes Federation. 2019. URL: https://www.diabetesatlas.org [accessed 2020-08-01]
2. American Diabetes Association. 1. Promoting health and reducing disparities in populations. Diabetes Care 2016 Dec 15;40(Supplement 1):S6-S10 [FREE Full text] [doi: 10.2337/dc17-s004]
3. Skovlund SE, Lichtenberg T, Hessler D, Ejskjaer N. Can the routine use of patient-reported outcome measures improve the delivery of person-centered diabetes care? A review of recent developments and a case study. Curr Diab Rep 2019 Aug 16;19(9):1-18 [FREE Full text] [doi: 10.1007/s11892-019-1190-x]
4. Weldring T, Smith SM. Article Commentary: Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs). Health Serv Ins 2013 Aug 04;6:HSI.S11093-HSI.S11068 [FREE Full text] [doi: 10.4137/hsi.s11093]
5. Boyer JG, Earp JAL. The development of an instrument for assessing the quality of life of people with diabetes. Med Care 1997 May;35(5):440-453 [FREE Full text] [doi: 10.1097/00005650-199705000-00003]
6. Welch GW, Jacobson AM, Polonsky WH. The Problem Areas in Diabetes Scale. An evaluation of its clinical utility. Diabetes Care 1997 May 01;20(5):760-766 [FREE Full text] [doi: 10.2337/diacare.20.5.760] [Medline: 9135939]
7. Lee E, Lee YW, Lee K, Kim YS, Nam M. Measurement of diabetes-related emotional distress using the Problem Areas in Diabetes scale: psychometric evaluations show that the short form is better than the full form. Health Qual Life Outcomes 2014 Oct 29;12(1):142 [FREE Full text] [doi: 10.1186/s12955-014-0142-z] [Medline: 25358396]

8. El Achhab Y, Nejjari C, Chikri M, Lyoussi B. Disease-specific health-related quality of life instruments among adults diabetic: A systematic review. Diabetes Res Clin Pract 2008 May;80(2):171-184. [doi: 10.1016/j.diabres.2007.12.020] [Medline: 18279993]
9. Lee J, Lee E, Chae D, Kim C. Patient-reported outcome measures for diabetes self-care: a systematic review of measurement properties. Int J Nurs Stud 2020 May;105:103498 [FREE Full text] [doi: 10.1016/j.ijnurstu.2019.103498] [Medline: 32203756]
10. Ortega-Avila A, Cervera-Garvi P, Ramos-Petersen L, Chicharro-Luna E, Gijon-Nogueron G. Patient-reported outcome measures for patients with diabetes mellitus associated with foot and ankle pathologies: a systematic review. J Clin Med 2019 Jan 27;8(2):146 [FREE Full text] [doi: 10.3390/jcm8020146] [Medline: 30691204]
11. Chen YT, Tan YZ, Cheen M, Wee H. Patient-reported outcome measures in registry-based studies of type 2 diabetes mellitus: a systematic review. Curr Diab Rep 2019 Nov 20;19(11):135 [FREE Full text] [doi: 10.1007/s11892-019-1265-8] [Medline: 31748944]
12. García-Inzunza JA, Valles-Medina AM, Muñoz FA, Delgadillo-Ramos G, Compean-Ortiz LG. Validity of the Mexican version of the combined Foot Care Confidence / Foot-Care Behavior scale for diabetes. Rev Panam Salud Publica 2015 Jul;38(1):35-41. [Medline: 26506319]
13. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
14. Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, et al. COSMIN Risk of Bias checklist for systematic reviews of patient-reported outcome measures. Qual Life Res 2018 May 19;27(5):1171-1179 [FREE Full text] [doi: 10.1007/s11136-017-1765-4] [Medline: 29260445]
15. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res 2018 May 12;27(5):1147-1157 [FREE Full text] [doi: 10.1007/s11136-018-1798-3] [Medline: 29435801]
16. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007 Jan;60(1):34-42 [FREE Full text] [doi: 10.1016/j.jclinepi.2006.03.012] [Medline: 17161752]
17. Measurement properties of patient reported outcome measures for Diabetes Mellitus: a systematic review. Open Science Framework. URL: https://osf.io/v2nwu [accessed 2021-07-12]
18. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res 2018 May;27(5):1147-1157 [FREE Full text] [doi: 10.1007/s11136-018-1798-3] [Medline: 29435801]
19. Terwee CB, Jansma EP, Riphagen II, de Vet HCW. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Qual Life Res 2009 Oct 27;18(8):1115-1123 [FREE Full text] [doi: 10.1007/s11136-009-9528-5] [Medline: 19711195]
20. COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN). 2010. URL: http://www.cosmin.nl [accessed 2020-04-04]
21. Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. Qual Life Res 2018 May 17;27(5):1159-1170 [FREE Full text] [doi: 10.1007/s11136-018-1829-0] [Medline: 29550964]
22. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 2010 Jul;63(7):737-745 [FREE Full text] [doi: 10.1016/j.jclinepi.2010.02.006] [Medline: 20494804]
23. Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, de Vet HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res 2012 May 6;21(4):651-657 [FREE Full text] [doi: 10.1007/s11136-011-9960-1] [Medline: 21732199]
24. Marcel Dijkers. Introducing GRADE: a systematic approach to rating evidence in systematic reviews and to guideline development. KT Update. 2013. URL: https://ktdrr.org/products/update/v1n5/dijkers_grade_ktupdatev1n5.pdf [accessed 2021-07-19]
25. Pibernik-Okanović M. Psychometric properties of the World Health Organisation quality of life questionnaire (WHOQOL-100) in diabetic patients in Croatia. Diabetes Res Clin Pract 2001 Feb;51(2):133-143. [doi: 10.1016/s0168-8227(00)00230-8]
26. Kolawole BA, Mosaku SK, Ikem RT. A comparison of two measures of quality of life of Nigerian clinic patients with type 2 diabetes mellitus. Afr Health Sci 2009 Sep;9(3):161-166 [FREE Full text] [Medline: 20589144]
27. Sreedevi A, Cherkil S, Kuttikattu DS, Kamalamma L, Oldenburg B. Validation of WHOQOL-BREF in Malayalam and determinants of quality of life among people with type 2 diabetes in Kerala, India. Asia Pac J Public Health 2016 Jan;28(1 Suppl):62S-69S [FREE Full text] [doi: 10.1177/1010539515605888] [Medline: 26419636]

28. Jahanlou AS, Karami NA. WHO quality of life-BREF 26 questionnaire: reliability and validity of the Persian version and compare it with Iranian diabetics quality of life questionnaire in diabetic patients. Prim Care Diabetes 2011 Jul;5(2):103-107. [doi: 10.1016/j.pcd.2011.02.001] [Medline: 21481660]
29. Reba K, Birhane BW, Gutema H. Validity and reliability of the Amharic version of the World Health Organization's Quality of Life Questionnaire (WHOQOL-BREF) in patients with diagnosed Type 2 diabetes in Felege Hiwot Referral Hospital, Ethiopia. J Diabetes Res 2019 May 06;2019:3513159-3513156 [FREE Full text] [doi: 10.1155/2019/3513159] [Medline: 31198788]
30. Maddigan SL, Feeny DH, Johnson JA. Construct validity of the RAND-12 and Health Utilities Index Mark 2 and 3 in type 2 diabetes. Qual Life Res 2004 Mar;13(2):435-448. [doi: 10.1023/b:qure.0000018497.06539.8f]
31. Morgan BS, Buscemi CP, Fajardo VP. Assessing instruments in a Cuban American population with type 2 diabetes mellitus. J Transcult Nurs 2004 Apr 29;15(2):139-146. [doi: 10.1177/1043659603262487] [Medline: 15070496]
32. Maddigan SL, Feeny DH, Majumdar SR, Farris KB, Johnson JA. Health Utilities Index mark 3 demonstrated construct validity in a population-based sample with type 2 diabetes. J Clin Epidemiol 2006 May;59(5):472-477. [doi: 10.1016/j.jclinepi.2005.09.010] [Medline: 16632135]
33. Matza LS, Boye KS, Yurgin N. Validation of two generic patient-reported outcome measures in patients with type 2 diabetes. Health Qual Life Outcomes 2007 Jul 31;5(1):47 [FREE Full text] [doi: 10.1186/1477-7525-5-47] [Medline: 17672906]
34. Lee WJ, Song K, Noh JH, Choi YJ, Jo M. Health-related quality of life using the EuroQol 5D questionnaire in Korean patients with type 2 diabetes. J Korean Med Sci 2012 Mar;27(3):255-260 [FREE Full text] [doi: 10.3346/jkms.2012.27.3.255] [Medline: 22379335]
35. Sayah FA, Qiu W, Xie F, Johnson JA. Comparative performance of the EQ-5D-5L and SF-6D index scores in adults with type 2 diabetes. Qual Life Res 2017 Aug 31;26(8):2057-2066. [doi: 10.1007/s11136-017-1559-8] [Medline: 28364183]
36. McClure NS, Sayah FA, Ohinmaa A, Johnson JA. Minimally important difference of the EQ-5D-5L Index Score in adults with type 2 diabetes. Value Health 2018 Sep;21(9):1090-1097 [FREE Full text] [doi: 10.1016/j.jval.2018.02.007] [Medline: 30224114]
37. Wang P, Luo N, Tai E, Thumboo J. The EQ-5D-5L is more discriminative than the EQ-5D-3L in patients with diabetes in Singapore. Value Health Reg Issues 2016 May;9:57-62. [doi: 10.1016/j.vhri.2015.11.003] [Medline: 27881260]
38. Konerding U, Elkhuizen SG, Faubel R, Forte P, Malmström T, Pavi E, et al. The validity of the EQ-5D-3L items: an investigation with type 2 diabetes patients from six European countries. Health Qual Life Outcomes 2014 Dec 05;12(1):181 [FREE Full text] [doi: 10.1186/s12955-014-0181-5] [Medline: 25479769]
39. Koh D, Abdullah AMKB, Wang P, Lin N, Luo N. Validation of Brunei's Malay EQ-5D Questionnaire in patients with type 2 diabetes. PLoS One 2016;11(11):e0165555 [FREE Full text] [doi: 10.1371/journal.pone.0165555] [Medline: 27835652]
40. Fan J, McCoy RG, Ziegenfuss JY, Smith SA, Borah BJ, Deming JR, et al. Evaluating the structure of the Patient Assessment of Chronic Illness Care (PACIC) survey from the patient's perspective. Ann Behav Med 2015 Feb 19;49(1):104-111. [doi: 10.1007/s12160-014-9638-3] [Medline: 25236671]
41. Aung E, Donald M, Coll JR, Williams GM, Doi SAR. Association between patient activation and patient-assessed quality of care in type 2 diabetes: results of a longitudinal study. Health Expect 2016 Apr 13;19(2):356-366 [FREE Full text] [doi: 10.1111/hex.12359] [Medline: 25773785]
42. Gugiu PC, Coryn C, Clark R, Kuehn A. Development and evaluation of the short version of the Patient Assessment of Chronic Illness Care instrument. Chronic Illn 2009 Dec 19;5(4):268-276. [doi: 10.1177/1742395309348072] [Medline: 19933249]
43. Gugiu C, Coryn C, Applegate B. Structure and measurement properties of the Patient Assessment of Chronic Illness Care instrument. J Eval Clin Pract 2010 Jun;16(3):509-516. [doi: 10.1111/j.1365-2753.2009.01151.x] [Medline: 20210824]
44. Abdul-Razak S, Ramli AS, Badlishah-Sham SF, Haniff J, EMPOWER-PAR Investigators. Validity and reliability of the patient assessment on chronic illness care (PACIC) questionnaire: the Malay version. BMC Fam Pract 2018 Jul 19;19(1):119 [FREE Full text] [doi: 10.1186/s12875-018-0807-5] [Medline: 30025525]
45. Hu J, Gruber KJ, Hsueh K. Psychometric properties of the Chinese version of the SF-36 in older adults with diabetes in Beijing, China. Diabetes Res Clin Pract 2010 Jun;88(3):273-281. [doi: 10.1016/j.diabres.2010.03.005] [Medline: 20392508]
46. Wan EYF, Choi EPH, Yu EYT, Chin WY, Fung CSC, Chan AKC, et al. Evaluation of the internal and external responsiveness of Short Form-12 Health Survey version 2 (SF-12v2) in patients with type 2 diabetes mellitus. Qual Life Res 2018 Sep 13;27(9):2459-2469. [doi: 10.1007/s11136-018-1908-2] [Medline: 29948606]
47. Chuayruang K, Sriratanaban J, Hiransuthikul N, Suwanwalaikorn S. Development of an instrument for patient-reported outcomes in Thai patients with type 2 diabetes mellitus (PRO-DM-Thai). Asian Biomed 2015;9:7-19. [doi: 10.5372/1905-7415.0901.363]
48. Cheng AY, Tsui EY, Hanley AJ, Zinman B. Developing a quality of life measure for Chinese patients with diabetes. Diabetes Research and Clinical Practice 1999 Dec;46(3):259-267. [doi: 10.1016/s0168-8227(99)00091-1]
49. Jin X, Liu GG, Gerstein HC, Levine MAH, Steeves K, Guan H, et al. Item reduction and validation of the Chinese version of diabetes quality-of-life measure (DQOL). Health Qual Life Outcomes 2018 Apr 27;16(1):78 [FREE Full text] [doi: 10.1186/s12955-018-0905-z] [Medline: 29703205]

50. Pakpour A, Saffari M, Burri A. Translation and validation of an Iranian version of the Diabetes Quality of Life measure. J Diabetes Investig 2012 Oct 18;3(5):471-478 [FREE Full text] [doi: 10.1111/j.2040-1124.2012.00217.x] [Medline: 24843609]
51. Yildirim A, Akinci F, Gozu H, Sargin H, Orbay E, Sargin M. Translation, cultural adaptation, cross-validation of the Turkish diabetes quality-of-life (DQOL) measure. Qual Life Res 2007 Jun 8;16(5):873-879. [doi: 10.1007/s11136-007-9172-x] [Medline: 17286193]
52. Bujang MA, Ismail M, Hatta NKBM, Othman SH, Baharum N, Lazim SSM. Validation of the Malay version of Diabetes Quality of Life (DQOL) Questionnaire for adult population with type 2 diabetes mellitus. Malays J Med Sci 2017 Aug 18;24(4):86-96 [FREE Full text] [doi: 10.21315/mjms2017.24.4.10] [Medline: 28951693]
53. Bujang MA, Adnan TH, Mohd Hatta NKB, Ismail M, Lim CJ. A revised version of Diabetes Quality of Life Instrument maintaining domains for satisfaction, impact, and worry. J Diabetes Res 2018 Jul 22;2018:5804687-5804610 [FREE Full text] [doi: 10.1155/2018/5804687] [Medline: 30327784]
54. Goh S, Rusli B, Khalid B. Development and validation of the Asian Diabetes Quality of Life (AsianDQOL) Questionnaire. Diabetes Res Clin Pract 2015 Jun;108(3):489-498 [FREE Full text] [doi: 10.1016/j.diabres.2015.02.009] [Medline: 25790899]
55. Dudzińska M, Tarach JS, Burroughs TE, Zwolak A, Matuszek B, Smoleń A, et al. Validation of the Polish version of Diabetes Quality of Life - Brief Clinical Inventory (DQL-BCI) among patients with type 2 diabetes. Arch Med Sci 2014 Oct 27;10(5):891-898 [FREE Full text] [doi: 10.5114/aoms.2014.46210] [Medline: 25395940]
56. Magwood GS, Jenkins C, Zapka J. Validation of diabetes health-related quality-of-life instruments using cognitive interviewing with older African Americans. J Nurs Meas 2009 Dec 01;17(3):195-220 [FREE Full text] [doi: 10.1891/1061-3749.17.3.195] [Medline: 20069949]
57. Nagpal J, Kumar A, Kakar S, Bhartia A. The development of 'Quality of Life Instrument for Indian Diabetes patients (QOLID): a validation and reliability study in middle and higher income groups. J Assoc Physicians India 2010 May;58:295-304. [Medline: 21117348]
58. Sato F, Mita T, Yamamoto R, Hirose T, Ito C, Tamura Y, et al. Reliability and validity of the Japanese version of the Diabetes Quality-Of-Life questionnaire for Japanese patients with type 2 diabetes mellitus. Diabetol Int 2013 Jul 2;5(1):21-29. [doi: 10.1007/s13340-013-0125-z]
59. Oobe M, Tanaka M, Fuchigami M, Sakata T. Preparation of a quality of life (QOL) questionnaire for patients with type II diabetes and prospects for its clinical application. Fukuoka Igaku Zasshi 2007 Oct;98(10):379-387. [Medline: 18046991]
60. Saffari M, Lin C, O'Garo K, Koenig HG, Sanaeinasab H, Pakpour AH. Psychometric properties of Persian Diabetes-Mellitus Specific Quality of Life (DMQoL) questionnaire in a population-based sample of Iranians. Int J Diabetes Dev Ctries 2018 Jun 4;39(1):218-227. [doi: 10.1007/s13410-018-0648-8]
61. Hasan SS, Ahmadi K, Santigo R, Ahmed SI. The validity of the Menopause-specific Quality of Life questionnaire in women with type 2 diabetes. Climacteric 2014 Aug 30;17(4):456-464. [doi: 10.3109/13697137.2013.864269] [Medline: 24228772]
62. Khader YS, Bataineh S, Batayha W. The Arabic version of Diabetes-39: psychometric properties and validation. Chronic Illn 2008 Dec 01;4(4):257-263. [doi: 10.1177/1742395308100647] [Medline: 19091934]
63. Queiroz FAD, Pace AE, Santos CBD. Cross-cultural adaptation and validation of the instrument Diabetes - 39 (D-39): brazilian version for type 2 diabetes mellitus patients - stage 1. Rev Lat Am Enfermagem 2009;17(5):708-715 [FREE Full text] [doi: 10.1590/s0104-11692009000500018] [Medline: 19967222]
64. Jannoo Z, Yap BW, Musa KI, Lazim MA, Hassali MA. An audit of diabetes-dependent quality of life in patients with type 2 diabetes mellitus in Malaysia. Qual Life Res 2015 Sep 24;24(9):2297-2302. [doi: 10.1007/s11136-015-0969-8] [Medline: 25800728]
65. Fung CSC, Wan EYF, Yu CLY, Wong CKH. Validity and reliability of the 19-item Audit of Diabetes-Dependent Quality of Life (ADDQoL-19) questionnaire in Chinese patients with type 2 diabetes mellitus in primary care. Qual Life Res 2016 Sep 15;25(9):2373-2378. [doi: 10.1007/s11136-016-1263-0] [Medline: 26980420]
66. Abbatecola AM, Spazzafumo L, Fabbietti P, Testa R, Rabini RA, Bonfigli AR, et al. Diabetes-related quality of life is enhanced by glycaemic improvement in older people. Diabet Med 2015 Feb 05;32(2):243-249. [doi: 10.1111/dme.12605] [Medline: 25307368]
67. Soon SS, Goh SY, Bee YM, Poon JL, Li SC, Thumboo J, et al. Audit of Diabetes-Dependent Quality of Life (ADDQoL) [Chinese Version for Singapore] questionnaire: reliability and validity among Singaporeans with type 2 diabetes mellitus. Appl Health Econ Health Policy 2010;8(4):239-249. [doi: 10.2165/11313920-000000000-00000] [Medline: 20578779]
68. Kong D, Ding Y, Zuo X, Su W, Xiu L, Lin M, et al. Adaptation of the Audit of Diabetes-Dependent Quality of Life questionnaire to people with diabetes in China. Diabetes Res Clin Pract 2011 Oct;94(1):45-52. [doi: 10.1016/j.diabres.2011.05.026] [Medline: 21683467]
69. Lemon SC, Rosal MC, Welch G. Measuring quality of life in low-income, Spanish-speaking Puerto Ricans with type 2 diabetes residing in the mainland U.S. Qual Life Res 2011 Nov 8;20(9):1507-1511. [doi: 10.1007/s11136-011-9871-1] [Medline: 21384266]
70. Demirci H, Cinar Y, Bayram N, Bilgel N. Quality of life in type II diabetic patients in primary health care. Dan Med J 2012 Oct;59(10):A4468 [FREE Full text] [Medline: 23158887]

71. Kamarul Imran M, Ismail AA, Naing L, Wan Mohamad WB. The reliability and validity of the Malay version of the 18-item audit of Diabetes Dependent Quality of Life (the Malay ADDQOL) questionnaire. Southeast Asian J Trop Med Public Health 2007 Mar;38(2):398-405. [Medline: 17539293]
72. Elasy TA, Samuel-Hodge CD, DeVellis RF, Skelly AH, Ammerman AS, Keyserling TC. Development of a health status measure for older African-American women with type 2 diabetes. Diabetes Care 2000 Mar 01;23(3):325-329. [doi: 10.2337/diacare.23.3.325]
73. Meadows KA, Abrams C, Sandbaek A. Adaptation of the Diabetes Health Profile (DHP-1) for use with patients with Type 2 diabetes mellitus: psychometric evaluation and cross-cultural comparison. Diabet Med 2000 Aug;17(8):572-580. [doi: 10.1046/j.1464-5491.2000.00322.x] [Medline: 11073178]
74. Mulhern B, Labeit A, Rowen D, Knowles E, Meadows K, Elliott J, et al. Developing preference-based measures for diabetes: DHP-3D and DHP-5D. Diabet Med 2017 Sep 01;34(9):1264-1275. [doi: 10.1111/dme.13377] [Medline: 28477411]
75. Li J, Li Z, Zhao W, Pan H, Halloran EJ. The reliability and validity of the diabetes care profile for Chinese populations. Eval Health Prof 2015 Jun 28;38(2):200-218. [doi: 10.1177/0163278714525628] [Medline: 24586109]
76. Li T, Lin C, Liu C, Li C, Lee Y. Validation of the Chinese version of the diabetes impact measurement scales amongst people suffering from diabetes. Qual Life Res 2006 Dec 7;15(10):1613-1619. [doi: 10.1007/s11136-006-0024-x] [Medline: 16826434]
77. Talbot F, Nouwen A, Gingras J, Gosselin M, Audet J. The assessment of diabetes-related cognitive and social factors: the Multidimensional Diabetes Questionnaire. J Behav Med 1997;20:291-312. [doi: 10.1037/t66308-000]
78. Pawar S, Thakurdesai P. Translation and validation of hindi version of the multidimensional questionnaire (MDQ) for quality of life assessment in type 2 diabetes mellitus patients in indian population. International Journal of Pharmaceutical Sciences and Research 2013;4:1827-1832 [FREE Full text]
79. Kolawole B, Abodunde O, Ikem R, Fabiyi A. A Test of the reliability and validity of a diabetes specific quality of life scale in a Nigerian hospital. Qual Life Res 2004 Sep;13(7):1287-1295. [doi: 10.1023/b:qure.0000037483.85545.5f]
80. Speight J, Khagram L, Davies M. Generic and diabetes-specific well-being in the AT.LANTUS Follow-on study: further psychometric validation of the W-BQ28 indicates its utility in research and clinical practice in Type 2 diabetes in the UK. Diabet Med 2012 Sep;29(9):e345-e353. [doi: 10.1111/j.1464-5491.2012.03702.x] [Medline: 22540936]
81. Hajos TRS, Pouwer F, Skovlund SE, Den Oudsten BL, Geelhoed-Duijvestijn PHLM, Tack CJ, et al. Psychometric and screening properties of the WHO-5 well-being index in adult outpatients with Type 1 or Type 2 diabetes mellitus. Diabet Med 2013 Feb 21;30(2):e63-e69. [doi: 10.1111/dme.12040] [Medline: 23072401]
82. Cichoń E, Kiejna A, Kokoszka A, Gondek T, Rajba B, Lloyd CE, et al. Validation of the Polish version of WHO-5 as a screening instrument for depression in adults with diabetes. Diabetes Res Clin Pract 2020 Jan;159:107970. [doi: 10.1016/j.diabres.2019.107970] [Medline: 31805355]
83. Rankin SH, Galbraith ME, Johnson S. Reliability and validity data for a Chinese translation of the Center for Epidemiological Studies-Depression. Psychol Rep 1993 Dec 06;73(3 Pt 2):1291-1298. [doi: 10.2466/pr0.1993.73.3f.1291] [Medline: 8115582]
84. McHale M, Hendrikz J, Dann F, Kenardy J. Screening for depression in patients with diabetes mellitus. Psychosom Med 2008;70:869-874. [doi: 10.1097/psy.0b013e318186dea9]
85. Zhang Y, Ting RZW, Lam MHB, Lam S, Yeung RO, Nan H, et al. Measuring depression with CES-D in Chinese patients with type 2 diabetes: the validity and its comparison to PHQ-9. BMC Psychiatry 2015 Aug 18;15(1):198 [FREE Full text] [doi: 10.1186/s12888-015-0580-0] [Medline: 26281832]
86. Carter J, Cogo-Moreira H, Herrmann N, Merino D, Yang P, Shah BR, et al. Validity of the Center for Epidemiological Studies Depression scale in Type 2 diabetes. J Psychosom Res 2016 Nov;90:91-97. [doi: 10.1016/j.jpsychores.2016.09.013] [Medline: 27772565]
87. Kokoszka A. Depression in Diabetes Self-Rating Scale: A screening tool. Diabetologia Doswiadczalna i Kliniczna 2008;8:A43-A47 [FREE Full text]
88. de Cock ES, Emons WH, Nefs G, Pop VJ, Pouwer F. Dimensionality and scale properties of the Edinburgh Depression Scale (EDS) in patients with type 2 diabetes mellitus: the DiaDDzoB study. BMC Psychiatry 2011 Aug 24;11(1). [doi: 10.1186/1471-244x-11-141]
89. Zauszniewski JA, Chung C, Krafcik K, Sousa VD. Psychometric testing of the Depressive Cognition Scale in women with type 2 diabetes. J Nurs Meas 2001 May 01;9(1):61-72. [doi: 10.1891/1061-3749.9.1.61]
90. Hsu L, Kao C, Wang M, Chang C, Tsai P. Psychometric testing of a Mandarin Chinese version of the Clinically Useful Depression Outcome Scale for patients diagnosed with type 2 diabetes mellitus. Int J Nurs Stud 2014 Dec;51(12):1595-1604. [doi: 10.1016/j.ijnurstu.2014.05.004] [Medline: 24951085]
91. Zhang Y, Ting R, Lam M, Lam J, Nan H, Yeung R, et al. Measuring depressive symptoms using the Patient Health Questionnaire-9 in Hong Kong Chinese subjects with type 2 diabetes. J Affect Disord 2013 Nov;151(2):660-666. [doi: 10.1016/j.jad.2013.07.014] [Medline: 23938133]
92. Janssen EPCJ, Köhler S, Stehouwer CDA, Schaper NC, Dagnelie PC, Sep SJS, et al. The Patient Health Questionnaire-9 as a screening tool for depression in individuals with type 2 diabetes mellitus: the Maastricht Study. J Am Geriatr Soc 2016 Nov 26;64(11):e201-e206. [doi: 10.1111/jgs.14388] [Medline: 27783384]

93. Udedi M, Muula AS, Stewart RC, Pence BW. The validity of the Patient Health Questionnaire-9 to screen for depression in patients with type-2 diabetes mellitus in non-communicable diseases clinics in Malawi. BMC Psychiatry 2019 Feb 27;19(1):81 [FREE Full text] [doi: 10.1186/s12888-019-2062-2] [Medline: 30813922]
94. Lupascu N, Timar B, Albai A, Roman D, Potre O, Timar R. Validation and cross-cultural adaptation of the depression Patient’s Health Questionnaire – 9 in the Romanian population of patients with Type 2 Diabetes Mellitus. DMSO 2019 May;12:841-849. [doi: 10.2147/dmso.s203099]
95. Ting RZW, Nan H, Yu MWM, Kong APS, Ma RCW, Wong RYM, et al. Diabetes-related distress and physical and psychological health in chinese type 2 diabetic patients. Diabetes Care 2011 May 11;34(5):1094-1096 [FREE Full text] [doi: 10.2337/dc10-1612] [Medline: 21398526]
96. Farm BAS, Perwitasari DA, Thobari JA, Cao Q, Krabbe PF, Postma MJ. Translation, revision, and validation of the Diabetes Distress Scale for Indonesian type 2 diabetic outpatients with various types of complications. Value Health Reg Issues 2017 May;12:63-73. [doi: 10.1016/j.vhri.2017.03.010] [Medline: 28648318]
97. Venkataraman K, Tan LSM, Bautista DCT, Griva K, Zuniga YLM, Amir M, et al. Psychometric properties of the Problem Areas in Diabetes (PAID) instrument in Singapore. PLoS One 2015 Sep 3;10(9):e0136759 [FREE Full text] [doi: 10.1371/journal.pone.0136759] [Medline: 26336088]
98. Jannoo Z, Yap BW, Khan NM, Farcomeni A. Assessing diabetes distress among type 2 diabetes mellitus in Malaysia using the problem areas in diabetes scale. Value Health Reg Issues 2019 May;18:159-164. [doi: 10.1016/j.vhri.2019.03.004] [Medline: 31082796]
99. Gross CC, Scain SF, Scheffel R, Gross JL, Hutz CS. Brazilian version of the Problem Areas in Diabetes Scale (B-PAID): validation and identification of individuals at high risk for emotional distress. Diabetes Res Clin Pract 2007 Jun;76(3):455-459. [doi: 10.1016/j.diabres.2006.09.022] [Medline: 17081645]
100. Eom YS, Park HS, Kim S, Yang SM, Nam MS, Lee HW, et al. Evaluation of stress in Korean patients with diabetes mellitus using the problem areas in diabetes-Korea questionnaire. Diabetes Metab J 2011 Apr;35(2):182-187 [FREE Full text] [doi: 10.4093/dmj.2011.35.2.182] [Medline: 21738901]
101. Lee E, Lee YW, Lee K, Kim YS, Nam M. Measurement of diabetes-related emotional distress using the Problem Areas in Diabetes scale: psychometric evaluations show that the short form is better than the full form. Health Qual Life Outcomes 2014 Oct 29;12(1):142 [FREE Full text] [doi: 10.1186/s12955-014-0142-z] [Medline: 25358396]
102. Huis In 't Veld EMJ, Makine C, Nouwen A, Karşıdağ, Kadıoğlu P, Karşıdağ K, et al. Validation of the Turkish version of the problem areas in diabetes scale. Cardiovasc Psychiatry Neurol 2011 Dec 13;2011:315068-315066 [FREE Full text] [doi: 10.1155/2011/315068] [Medline: 22195273]
103. Papathanasiou A, Koutsovasilis A, Shea S, Philalithis A, Papavasiliou S, Melidonis A, et al. The Problem Areas in Diabetes (PAID) scale: psychometric evaluation survey in a Greek sample with type 2 diabetes. J Psychiatr Ment Health Nurs 2014 May 17;21(4):345-353. [doi: 10.1111/j.1365-2850.2012.01875.x] [Medline: 22340071]
104. Siaw M, Tai B, Lee J. Psychometric properties of the Chinese version of Problem Areas In Diabetes Scale (Sg-Paid-C) among high-risk polypharmacy patients with uncontrolled type 2 diabetes In Singapore. Value in Health 2016 Nov;19(7):A901. [doi: 10.1016/j.jval.2016.08.207]
105. Welch G, Schwartz CE, Santiago-Kelly P, Garb J, Shayne R, Bode R. Disease-related emotional distress of Hispanic and non-Hispanic type 2 diabetes patients. Ethn Dis 2007;17(3):541-547. [Medline: 17985511]
106. Arzaghi SM, Mahjouri MY, Heshmat R, Khashayar P, Larijani B. Psychometric properties of the Iranian version of the Problem Areas in Diabetes scale (IR-PAID-20). J Diabetes Metab Disord 2011;10:16 [FREE Full text]
107. Bijl J, Poelgeest-Eeltink A, Shortridge-Baggett L. The psychometric properties of the diabetes management self-efficacy scale for patients with type 2 diabetes mellitus. J Adv Nurs 1999 Aug;30(2):352-359. [doi: 10.1046/j.1365-2648.1999.01077.x] [Medline: 10457237]
108. Azami G, Kim Lam S, Shariff-Ghazali S, Mohd Said S, Aazami S, Mozafari M. Predictors of diabetes self-management behaviors in type 2 diabetes patients. IJDO 2021 Jan 16. [doi: 10.18502/ijdo.v12i4.5180]
109. Sturt J, Hearnshaw H, Wakelin M. Validity and reliability of the DMSES UK: a measure of self-efficacy for type 2 diabetes self-management. Prim Health Care 2010 Jun 22;11(04):374-381. [doi: 10.1017/s1463423610000101]
110. Lee E, van der Bijl J, Shortridge-Baggett LM, Han SJ, Moon SH. Psychometric properties of the Diabetes Management Self-Efficacy Scale in Korean patients with type 2 diabetes. Int J Endocrinol 2015;2015:780701 [FREE Full text] [doi: 10.1155/2015/780701] [Medline: 26089892]
111. Fappa E, Efthymiou V, Landis G, Rentoumis A, Doupis J. Validation of the Greek version of the Diabetes Management Self-Efficacy Scale (GR-DMSES). Adv Ther 2016 Jan 21;33(1):82-95. [doi: 10.1007/s12325-015-0278-1]
112. Pace AE, Gomes LC, Bertolin DC, Loureiro HMAM, Bijl JVD, Shortridge-Baggett LM. Adaptation and validation of the Diabetes Management Self-Efficacy Scale to Brazilian Portuguese. Rev Latino-Am Enfermagem 2017;25. [doi: 10.1590/1518-8345.1543.2861]
113. Messina R, Rucci P, Sturt J, Mancini T, Fantini MP. Assessing self-efficacy in type 2 diabetes management: validation of the Italian version of the Diabetes Management Self-Efficacy Scale (IT-DMSES). Health Qual Life Outcomes 2018 Apr 23;16(1):71 [FREE Full text] [doi: 10.1186/s12955-018-0901-3] [Medline: 29685153]

114. Thojampa S. Psychometric evaluation of the Thai translation of the Self-Efficacy for Diabetes Scale in delaying the progress of diabetic nephropathy in adults with T2DM. Int J Afr Nurs Sci 2018;9:1-3. [doi: 10.1016/j.ijans.2018.05.006]
115. Zheng J, Wang Y, Ye X, Xiao L, Ye J, Li X, et al. Validation of diabetes medication self-efficacy scale in Chinese with type 2 diabetes. Patient Prefer Adherence 2018 Nov;Volume 12:2517-2525. [doi: 10.2147/ppa.s170144]
116. Sousa V, Hartman S, Miller E, Carroll M. New measures of diabetes self-care agency, diabetes self-efficacy, and diabetes self-management for insulin-treated individuals with type 2 diabetes. J Clin Nurs 2009 May;18(9):1305-1312. [doi: 10.1111/j.1365-2702.2008.02729.x] [Medline: 19413558]
117. Chang SJ, Song M, Im E. Psychometric evaluation of the Korean version of the Diabetes Self-efficacy Scale among South Korean older adults with type 2 diabetes. J Clin Nurs 2014 Aug 22;23(15-16):2121-2130. [doi: 10.1111/jocn.12133] [Medline: 23876167]
118. van der Heijden MMP, Pouwer F, Pop VJM. Psychometric properties of the Exercise Self-efficacy Scale in Dutch Primary care patients with type 2 diabetes mellitus. Int J Behav Med 2014 Apr 3;21(2):394-401. [doi: 10.1007/s12529-013-9308-z] [Medline: 23550033]
119. Fain JA. Psychometric properties of the Spanish Version of the Diabetes Self-management Assessment Report Tool. Diabetes Educ 2007 Dec 07;33(5):827-832. [doi: 10.1177/0145721707308355] [Medline: 17925586]
120. Toobert DJ, Hampson SE, Glasgow RE. The summary of diabetes self-care activities measure: results from 7 studies and a revised scale. Diabetes Care 2000 Jul 01;23(7):943-950 [FREE Full text] [doi: 10.2337/diacare.23.7.943] [Medline: 10895844]
121. Kav S, Akman A, Dogan N, Tarakci Z, Bulut Y, Hanoglu Z. Turkish validity and reliability of the summary of diabetes self-care activities measure for patients with type 2 diabetes mellitus. J Clin Nurs 2010 Oct;19(19-20):2933-2935. [doi: 10.1111/j.1365-2702.2010.03329.x] [Medline: 20846234]
122. Kamradt M, Bozorgmehr K, Krisam J, Freund T, Kiel M, Qreini M, et al. Assessing self-management in patients with diabetes mellitus type 2 in Germany: validation of a German version of the Summary of Diabetes Self-Care Activities measure (SDSCA-G). Health Qual Life Outcomes 2014 Dec 18;12(1):185 [FREE Full text] [doi: 10.1186/s12955-014-0185-1] [Medline: 25519204]
123. Adarmouch L, Sebbani M, Elyacoubi A, Amine M. Psychometric properties of a Moroccan version of the Summary of Diabetes Self-Care Activities Measure. J Diabetes Res 2016;2016:5479216-5479216 [FREE Full text] [doi: 10.1155/2016/5479216] [Medline: 27019853]
124. Sukkarieh-Haraty O, Howard E. Psychometric properties of the Arabic version of the Summary of Diabetes Self-Care Activities Instrument. Res Theory Nurs Pract 2016 Jan 01;30(1):60-69. [doi: 10.1891/1541-6577.30.1.60] [Medline: 27025000]
125. Choi EJ, Nam M, Kim SH, Park CG, Toobert DJ, Yoo JS, et al. Psychometric properties of a Korean version of the summary of diabetes self-care activities measure. Int J Nurs Stud 2011 Mar;48(3):333-337. [doi: 10.1016/j.ijnurstu.2010.08.007] [Medline: 20950807]
126. Mendonça SCBD, Zanetti ML, Sawada NO, Barreto IDDC, Andrade JSD, Otero LM. Construction and validation of the Self-care Assessment Instrument for patients with type 2 diabetes mellitus. Rev Lat Am Enfermagem 2017 Jun 05;25:e2890 [FREE Full text] [doi: 10.1590/1518-8345.1533.2890] [Medline: 28591298]
127. Khagram L, Martin CR, Davies MJ, Speight J. Psychometric validation of the Self-Care Inventory-Revised (SCI-R) in UK adults with type 2 diabetes using data from the AT.LANTUS Follow-on study. Health Qual Life Outcomes 2013 Feb 26;11(1):24 [FREE Full text] [doi: 10.1186/1477-7525-11-24] [Medline: 23443007]
128. García AA. The Diabetes Symptom Self-Care Inventory: development and psychometric testing with Mexican Americans. J Pain Symptom Manage 2011 Apr;41(4):715-727 [FREE Full text] [doi: 10.1016/j.jpainsymman.2010.06.018] [Medline: 21276705]
129. McCaskill GM, Bolland KA, Burgio KL, Leeper J. Development and validation of a diabetes self-management instrument for older African-Americans. Soc Work Health Care 2016 Apr 04;55(5):381-394. [doi: 10.1080/00981389.2015.1129012] [Medline: 27045578]
130. Nakawatase Y, Taru C, Tsutou A, Shiotani H, Kido Y, Ohara T, et al. Development of an evaluation scale for self-management behavior related to physical activity of type 2 diabetic patients. Diabetes Care 2007 Nov 20;30(11):2843-2848. [doi: 10.2337/dc07-0685] [Medline: 17644618]
131. Thojampa S, Mawn B. Psychometric evaluation of the Thai translation of the Diabetes Self-management Questionnaire in type 2 diabetes. Int J Nurs Sci 2017 Jul 10;4(3):236-238 [FREE Full text] [doi: 10.1016/j.ijnss.2017.06.006] [Medline: 31406746]
132. Bukhsh A, Lee SWH, Pusparajah P, Schmitt A, Khan TM. Psychometric properties of the Diabetes Self-Management Questionnaire (DSMQ) in Urdu. Health Qual Life Outcomes 2017 Oct 12;15(1):200 [FREE Full text] [doi: 10.1186/s12955-017-0776-8] [Medline: 29025432]
133. Dao-Tran T, Anderson DJ, Chang AM, Seib C, Hurst C. Vietnamese version of Diabetes Self-Management Instrument: development and psychometric testing. Res Nurs Health 2017 Apr 09;40(2):177-184. [doi: 10.1002/nur.21777] [Medline: 27933635]

134. Lee C, Lin C, Anderson R. Psychometric evaluation of the Diabetes Self-Management Instrument Short Form (DSMI-20). Appl Nurs Res 2016 Feb;29:83-88. [doi: 10.1016/j.apnr.2015.04.013] [Medline: 26856494]
135. Seo K, Song M, Choi S, Kim S, Chang SJ. Development of a scale to measure diabetes self-management behaviors among older Koreans with type 2 diabetes, based on the seven domains identified by the American Association of Diabetes Educators. Jpn J Nurs Sci 2017 Apr 13;14(2):161-170. [doi: 10.1111/jjns.12145] [Medline: 27734627]
136. Peyrot M, Bushnell DM, Best JH, Martin ML, Cameron A, Patrick DL. Development and validation of the self-management profile for type 2 diabetes (SMP-T2D). Health Qual Life Outcomes 2012 Oct 05;10(1):125 [FREE Full text] [doi: 10.1186/1477-7525-10-125] [Medline: 23039868]
137. Laranjo L, Dias V, Nunes C, Paiva D, Mahoney B. Translation and validation of the Patient Activation Measure in Portuguese people with type 2 diabetes Mellitus. Acta Med Port 2018 Aug 31;31(7-8):382-390 [FREE Full text] [doi: 10.20344/amp.9072] [Medline: 30189166]
138. Chernyak N, Jülich F, Kasperidus J, Stephan A, Begun A, Kaltheuner M, et al. Time cost of diabetes: development of a questionnaire to assess time spent on diabetes self-care. J Diabetes Complications 2017 Jan;31(1):260-266. [doi: 10.1016/j.jdiacomp.2016.06.016] [Medline: 27411888]
139. Manit A, Tuicomepee A, Jiamjarasrangsi W, Taneepanichskul S. Development of needs and resources for self-management assessment instrument in Thais with type 2 diabetes: cross-cultural adaptation. J Med Assoc Thai 2011 Nov;94(11):1304-1313. [Medline: 22256469]
140. Mahjouri MY, Arzaghi SM, Heshmat R, Khashayar P, Esfahani EN, Larijani B. Psychometric properties of the Iranian version of Diabetes Empowerment Scale (IR-DES-28). J Diabetes Metab Disord 2012 Aug 02;11(1):4 [FREE Full text] [doi: 10.1186/2251-6581-11-4] [Medline: 23497711]
141. Hara Y, Iwashita S, Okada A, Tajiri Y, Nakayama H, Kato T, et al. Development of a novel, short, self-completed questionnaire on empowerment for patients with type 2 diabetes mellitus and an analysis of factors affecting patient empowerment. Biopsychosoc Med 2014;8(1):19 [FREE Full text] [doi: 10.1186/1751-0759-8-19] [Medline: 25183994]
142. Ching SM, Yee A, Lee PY, Ramachandran V, Shum KM, Ismael NF, et al. Psychometric properties of the Malay version of the diabetes empowerment scale among hospital Serdang type 2 diabetes mellitus patients using exploratory factor analysis. Health Qual Life Outcomes 2020 Feb 07;18(1):23 [FREE Full text] [doi: 10.1186/s12955-020-1280-0] [Medline: 32033609]
143. Chaves FF, Reis IA, Pagano AS, Torres HDC. Translation, cross-cultural adaptation and validation of the Diabetes Empowerment Scale – Short Form. Rev Saúde Pública 2017;51. [doi: 10.1590/s1518-8787.2017051006336]
144. Sousa M, Almeida M, Loureiro H, Martins T. Study of the psychometric properties of the Diabetes Empowerment Scale Short Form (DES-SF). Port J Public Health 2019 Nov 29;37(2-3):66-72. [doi: 10.1159/000504629]
145. Saffari M, Karimi T, Koenig HG, Al-Zaben F. Psychometric evaluation of the Persian version of the Type 2 Diabetes and Health Promotion Scale (T2DHPS): a diabetes-specific measure of lifestyle. Scand J Caring Sci 2015 Sep 18;29(3):603-612. [doi: 10.1111/scs.12181] [Medline: 25236973]
146. Yildiz E, Kavuran E. The validity and reliability of the type 2 diabetes and health promotion scale Turkish version: a methodological study. Scand J Caring Sci 2018 Mar 22;32(1):417-421. [doi: 10.1111/scs.12483] [Medline: 28833366]
147. Wang R, Lin L, Cheng C, Hsu M, Kao C. The psychometric testing of the diabetes health promotion self-care scale. J Nurs Res 2012 Jun;20(2):122-130. [doi: 10.1097/jnr.0b013e318254eb47] [Medline: 22592107]
148. Akohoue SA, Wallston KA, Schlundt DG, Rothman RL. Psychometric evaluation of the short version of the Personal Diabetes Questionnaire to assess dietary behaviors and exercise in patients with type 2 diabetes. Eating Behaviors 2017 Aug;26:182-188. [doi: 10.1016/j.eatbeh.2017.04.002]
149. Cheng L, Leung DYP, Wu Y, Sit JWH, Yang M, Li X. Psychometric properties of the Modified Personal Diabetes Questionnaire among Chinese patients with type 2 diabetes. Eval Health Prof 2018 Mar 20;41(1):3-24. [doi: 10.1177/0163278716664393] [Medline: 27649714]
150. Lewis K, Jennings A, Ward J, Bradley C. Health belief scales developed specifically for people with tablet-treated type 2 diabetes. Diabet Med 1990 Feb;7(2):148-155. [doi: 10.1111/j.1464-5491.1990.tb01350.x] [Medline: 2137756]
151. Brown SA, Becker HA, Garcia AA, Barton SA, Hanis CL. Measuring health beliefs in Spanish-speaking Mexican Americans with type 2 diabetes: adapting an existing instrument. Res Nurs Health 2002 Apr 14;25(2):145-158. [doi: 10.1002/nur.10023] [Medline: 11933008]
152. Kartal A, Ozsoy SA. Validity and reliability study of the Turkish version of Health Belief Model Scale in diabetic patients. Int J Nurs Stud 2007 Nov;44(8):1447-1458. [doi: 10.1016/j.ijnurstu.2007.06.004] [Medline: 17655849]
153. Lujan J. Testing the diabetes knowledge and health beliefs of Mexican Americans on the Texas–Mexico border. Hisp Hlth Care Int 2008 Mar 01;6(1):9-20. [doi: 10.1891/1540-4153.6.1.9]
154. Gerber B, Smith Jr EV, Girotti M, Pelaez L, Lawless K, Smolin L, et al. Using Rasch measurement to investigate the cross-form equivalence and clinical utility of Spanish and English versions of a diabetes questionnaire: a pilot study. J Appl Meas 2002;3(3):243-271. [Medline: 12147912]
155. Padhy M, Padiri RA, Hariharan M, Rana S. Diabetes Mellitus Knowledge Test: development, psychometric evaluation, and establishing norms for Indian population. Int J Diabetes Dev Ctries 2018 May 7;39(1):206-217. [doi: 10.1007/s13410-018-0644-z]

156. Matin H, Nadrian H, Sarbakhsh P, Shaghaghi A. Factorial structure and psychometric analysis of the Persian version of Perceived Competence Scale for Diabetes (PCSD-P). Behav Sci (Basel) 2019 May 07;9(5):50 [FREE Full text] [doi: 10.3390/bs9050050] [Medline: 31067721]
157. Miller CK, Achterberg CL. Reliability and validity of a nutrition and food-label knowledge test for women with Type 2 diabetes mellitus. J Nutr Educ 2000 Jan;32(1):43-48. [doi: 10.1016/s0022-3182(00)70509-7]
158. Miller C, Edwards L. Development and validation of a shelf inventory to evaluate household food purchases among older adults with diabetes mellitus. J Nutr Educ Behav 2002 Sep;34(5):261-267. [doi: 10.1016/s1499-4046(06)60104-8]
159. Samuel-Hodge CD, DeVellis RF, Ammerman A, Keyserling TC, Elasy TA. Reliability and validity of a measure of perceived diabetes and dietary competence in African American women with type 2 diabetes. Diabetes Educ 2002 Sep 04;28(6):979-988. [doi: 10.1177/014572170202800612] [Medline: 12526638]
160. Han CY, Zheng X, Lee LF, Chan CG, Lee YQ, Zailani NA, et al. Development of a diabetes-related nutrition knowledge questionnaire for individuals with type 2 diabetes mellitus in Singapore. Nutr Diet 2019 Nov 31;76(5):567-573. [doi: 10.1111/1747-0080.12513] [Medline: 30706657]
161. Finbråten HS, Guttersrud, Nordström G, Pettersen KS, Trollvik A, Wilde-Larsson B. Validating the Functional, Communicative, and Critical Health Literacy Scale using Rasch modeling and confirmatory factor analysis. J Nurs Meas 2018 Aug 01;26(2):341-363. [doi: 10.1891/1061-3749.26.2.341] [Medline: 30567948]
162. Kang SJ, Sim KH, Song BR, Park J, Chang SJ, Park C, et al. Validation of the health literacy scale for diabetes as a criterion-referenced test with standard setting procedures. Patient Educ Couns 2018 Aug;101(8):1468-1476. [doi: 10.1016/j.pec.2018.03.013] [Medline: 29598965]
163. Lee E, Lee YW. First-order vs. second-order structural validity of the Health Literacy Scale in patients with diabetes. Scand J Caring Sci 2018 Mar 03;32(1):441-447. [doi: 10.1111/scs.12460] [Medline: 28771769]
164. Luo H, Patil SP, Wu Q, Bell RA, Cummings DM, Adams AD, et al. Validation of a combined health literacy and numeracy instrument for patients with type 2 diabetes. Patient Educ Couns 2018 Oct;101(10):1846-1851. [doi: 10.1016/j.pec.2018.05.017] [Medline: 29805071]
165. Ashok KM, Shanmugasundaram P. Development of a validated questionnaire to assess knowledge and awareness among uncontrolled diabetic patients toward diabetes mellitus. Drug Invent Today 2019;11:801-806 [FREE Full text]
166. Ashok Kumar M, Shanmugasundaram P. A validation study of questionnaire towards mobile based health applications in uncontrolled diabetic population of India (South). Diabetes Metab Syndr 2019 May;13(3):2106-2110. [doi: 10.1016/j.dsx.2019.04.037] [Medline: 31235144]
167. Finbråten HS, Pettersen KS, Wilde-Larsson B, Nordström G, Trollvik A, Guttersrud. Validating the European Health Literacy Survey Questionnaire in people with type 2 diabetes: latent trait analyses applying multidimensional Rasch modelling and confirmatory factor analysis. J Adv Nurs 2017 Nov 29;73(11):2730-2744. [doi: 10.1111/jan.13342] [Medline: 28543754]
168. Moock J, Hessel F, Ziegeler D, Kubiak T, Kohlmann T. Development and testing of the Insulin Treatment Experience Questionnaire (ITEQ). Patient 2010 Mar 01;3(1):45-58. [doi: 10.2165/11319510-000000000-00000] [Medline: 22273275]
169. Ishii H, Shin H, Tosaki T, Haga T, Nakajima Y, Shiraiwa T, et al. Reproducibility and validity of a questionnaire measuring treatment burden on patients with type 2 diabetes: Diabetic Treatment Burden Questionnaire (DTBQ). Diabetes Ther 2018 Jun 29;9(3):1001-1019 [FREE Full text] [doi: 10.1007/s13300-018-0414-4] [Medline: 29600503]
170. Rubin RR, Peyrot M. Psychometric properties of an instrument for assessing the experience of patients treated with inhaled insulin: the inhaled insulin treatment questionnaire (IITQ). Health Qual Life Outcomes 2010 Mar 24;8(1):32 [FREE Full text] [doi: 10.1186/1477-7525-8-32] [Medline: 20334647]
171. Kontodimopoulos N, Arvanitaki E, Aletras VH, Niakas D. Psychometric properties of the Greek Diabetes Treatment Satisfaction Questionnaire. Health Qual Life Outcomes 2012 Feb 01;10(1):17 [FREE Full text] [doi: 10.1186/1477-7525-10-17] [Medline: 22296783]
172. Brod M, Christensen T, Kongsø JH, Bushnell DM. Examining and interpreting responsiveness of the Diabetes Medication Satisfaction measure. J Med Econ 2009 Oct 08;12(4):309-316. [doi: 10.3111/13696990903337017] [Medline: 19811109]
173. Snoek FJ, Skovlund SE, Pouwer F. Development and validation of the insulin treatment appraisal scale (ITAS) in patients with type 2 diabetes. Health Qual Life Outcomes 2007 Dec 20;5:69 [FREE Full text] [doi: 10.1186/1477-7525-5-69] [Medline: 18096074]
174. Lee KP. Translation and validation of the Insulin Treatment Appraisal Scale in Hong Kong primary care patients. J Diabetes Investig 2018 Mar;9(2):311-320 [FREE Full text] [doi: 10.1111/jdi.12704] [Medline: 28626953]
175. Saritas SC, Erci B, Sahin I, Timocin E. Validity study of the Turkish version of the barriers to insulin treatment questionnaire. Int J Diabetes Dev Ctries 2018 Aug 9;39(2):247-253. [doi: 10.1007/s13410-018-0670-x]
176. Fu SN, Chin WY, Wong CKH, Yeung VTF, Yiu MP, Tsui HY, et al. Development and validation of the Chinese Attitudes to Starting Insulin Questionnaire (Ch-ASIQ) for primary care patients with type 2 diabetes. PLoS One 2013 Nov 13;8(11):e78933 [FREE Full text] [doi: 10.1371/journal.pone.0078933] [Medline: 24236071]
177. Osborn CY, Gonzalez JS. Measuring insulin adherence among adults with type 2 diabetes. J Behav Med 2016 Aug 9;39(4):633-641 [FREE Full text] [doi: 10.1007/s10865-016-9741-y] [Medline: 27062271]

178. Pedrosa RBDS, Trevisan DD, Nascimento RAD, São-João TM, Lima MHDM, Rodrigues RCM. Psychometric performance of the Brazilian version the “Insulin Management Diabetes Self-Efficacy Scale” for patients with type 2 diabetes mellitus. Article in Portuguese. Medicina (Ribeirao Preto) 2018 Nov 22;51(2):121-130. [doi: 10.11606/issn.2176-7262.v51i2p121-130]
179. Anderson R, Skovlund S, Marrero D, Levine D, Meadows K, Brod M, et al. Development and validation of the Insulin Treatment Satisfaction Questionnaire. Clin Ther 2004;26:565-578. [doi: 10.1037/t68803-000]
180. Ishii H, Oda E. Reproducibility and validity of a satisfaction questionnaire on hypoglycemic agents: the Oral Hypoglycemic Agent Questionnaire (OHA-Q). Diabetol Int 2012 Aug 1;3(3):152-163. [doi: 10.1007/s13340-012-0074-y]
181. Peyrot M, Harshaw Q, Shillington A, Xu Y, Rubin R. Validation of a tool to assess medication treatment satisfaction in patients with Type 2 diabetes: the Diabetes Medication System Rating Questionnaire (DMSRQ). Diabet Med 2012;29:1060-1066. [doi: 10.1111/j.1464-5491.2011.03538.x]
182. Demirtaş A, Akbayrak N. Development of an assessment scale for treatment compliance in type 2 Diabetes Mellitus in Turkish population: psychometric evaluation. Int J Nurs Sci 2017 Jul 10;4(3):244-251 [FREE Full text] [doi: 10.1016/j.ijnss.2017.06.002] [Medline: 31406748]
183. Sakthong P, Chabunthom R, Charoenvisuthiwongs R. Psychometric properties of the Thai version of the 8-item Morisky Medication Adherence Scale in patients with type 2 diabetes. Ann Pharmacother 2009 Apr 14;43(5):950-957. [doi: 10.1345/aph.1l453]
184. Wang Y, Lee J, Toh M, Tang W, Ko Y. Validity and reliability of a self-reported measure of medication adherence in patients with Type 2 diabetes mellitus in Singapore. Diabet Med 2012 Sep;29(9):e338-e344. [doi: 10.1111/j.1464-5491.2012.03733.x] [Medline: 22672497]
185. Lee W, Ahn J, Kim J, Hong Y, Hong SK, Kim YT, et al. Reliability and validity of a self-reported measure of medication adherence in patients with type 2 diabetes mellitus in Korea. J Int Med Res 2013 Aug 16;41(4):1098-1110 [FREE Full text] [doi: 10.1177/0300060513484433] [Medline: 23860015]
186. Wang J, Bian R, Mo Y. Validation of the Chinese version of the eight-item Morisky medication adherence scale in patients with type 2 diabetes mellitus. J Clin Gerontol Geriatr 2013 Dec;4(4):119-122. [doi: 10.1016/j.jcgg.2013.06.002]
187. Zongo A, Guénette L, Moisan J, Grégoire JP. Predictive validity of self-reported measures of adherence to noninsulin antidiabetes medication against control of glycated hemoglobin levels. Can J Diabetes 2016 Feb;40(1):58-65. [doi: 10.1016/j.jcjd.2015.06.008] [Medline: 26507401]
188. Zongo A, Guénette L, Moisan J, Guillaumie L, Lauzier S, Grégoire JP. Revisiting the internal consistency and factorial validity of the 8-item Morisky Medication Adherence Scale. SAGE Open Med 2016;4:2050312116674850 [FREE Full text] [doi: 10.1177/2050312116674850] [Medline: 27895914]
189. Kristina SA, Putri LR, Riani DA, Ikawati Z, Endarti D. Validity of self-reported measure of medication adherence among diabetic patients in Indonesia. Int Res J Pharm 2019 Aug 2;10(7):144-148. [doi: 10.7897/2230-8407.1007234]
190. Prado-Aguilar CA, Martínez YV, Segovia-Bernal Y, Reyes-Martínez R, Arias-Ulloa R. Performance of two questionnaires to measure treatment adherence in patients with Type-2 diabetes. BMC Public Health 2009 Jan 26;9:38 [FREE Full text] [doi: 10.1186/1471-2458-9-38] [Medline: 19171059]
191. Fall E, Gauchet A, Izaute M, Horne R, Chakroun N. Validation of the French version of the Beliefs about Medicines Questionnaire (BMQ) among diabetes and HIV patients. Eur Rev Appl Psychol 2014 Nov;64(6):335-343. [doi: 10.1016/j.erap.2014.08.005]
192. Chung WW, Chua SS, Lai PSM, Morisky DE. The Malaysian Medication Adherence Scale (MALMAS): concurrent validity using a clinical measure among people with type 2 diabetes in Malaysia. PLoS One 2015 Apr 24;10(4):e0124275 [FREE Full text] [doi: 10.1371/journal.pone.0124275] [Medline: 25909363]
193. Boas LCG, Lima MLSAPD, Pace AE. Adherence to treatment for diabetes mellitus: validation of instruments for oral antidiabetics and insulin. Rev Lat Am Enfermagem 2014 Jan;22(1):11-18 [FREE Full text] [doi: 10.1590/0104-1169.3155.2386] [Medline: 24553698]
194. Kim C, Park E, Schlenk EA, Kim M, Kim DJ. Psychometric evaluation of a Korean version of the Adherence to Refills and Medications Scale (ARMS) in adults with type 2 diabetes. Diabetes Educ 2016 Apr 22;42(2):188-198. [doi: 10.1177/0145721716632062] [Medline: 26902527]
195. Peyrot M, Xu Y, Rubin RR. Development and validation of the Diabetes Medication System Rating Questionnaire-Short Form. Diabet Med 2014 Oct 18;31(10):1237-1244 [FREE Full text] [doi: 10.1111/dme.12453] [Medline: 24673614]
196. Kawata AK, Wilson H, Ong SH, Kulich K, Coyne K. Development and psychometric evaluation of the Hypoglycemia Perspectives Questionnaire in patients with type 2 diabetes mellitus. Patient 2016 Oct 12;9(5):395-407. [doi: 10.1007/s40271-016-0163-2] [Medline: 26970976]
197. Çinar D, Yava A. Validity and reliability of functional assessment of chronic illness treatment-fatigue scale in Turkish patients with type 2 diabetes. Endocrinol Diabetes Nutr (Engl Ed) 2018 Aug;65(7):409-417. [doi: 10.1016/j.endinu.2018.01.010] [Medline: 29685730]
198. de Mesa U, Anonuevo-Cruz MC, Nicodemus N, Reyes N. Development and validation of a questionnaire evaluating impaired hypoglycemia awareness among adult filipino patients with type 2 diabetes mellitus. J ASEAN Fed Endocr Soc 2017 Sep 30;32(2):158-164 [FREE Full text] [doi: 10.15605/jafes.032.02.10] [Medline: 33442100]

199. Liu Y, Xiong S, Sang M, Li Y, Anarte Ortiz MT, Xing Q, et al. Reliability and validity of the Chinese version of the new fear of hypoglycemia scale: FH-15. Int J Nurs Sci 2018 Oct 10;5(4):343-351 [FREE Full text] [doi: 10.1016/j.ijnss.2018.09.008] [Medline: 31406846]
200. Lee E, Lee K, Song R, Snoek FJ, Moon S. Psychometric evaluation of the Korean version of the Diabetes Symptom Checklist-Revised (DSC-R) for patients with type 2 diabetes. Health Qual Life Outcomes 2014 May 19;12(1):77 [FREE Full text] [doi: 10.1186/1477-7525-12-77] [Medline: 24885358]
201. Arbuckle RA, Humphrey L, Vardeva K, Arondekar B, Danten-Viala M, Scott JA, et al. Psychometric evaluation of the Diabetes Symptom Checklist-Revised (DSC-R)--a measure of symptom distress. Value Health 2009 Nov;12(8):1168-1175 [FREE Full text] [doi: 10.1111/j.1524-4733.2009.00571.x] [Medline: 19558371]
202. Naegeli AN, Hayes C. A psychometric evaluation of the Diabetes Symptom Checklist-Revised (DSC-R) cognitive distress, fatigue, hyperglycemia, and hypoglycemia subscales in patients with type 1 and type 2 diabetes. DMSO 2010 Mar;Volume 3:27-30. [doi: 10.2147/dmso.s9465]
203. Sato E, Suzukamo Y, Miyashita M, Kazuma K. Development of a diabetes diet-related quality-of-life scale. Diabetes Care 2004 Jun 25;27(6):1271-1275. [doi: 10.2337/diacare.27.6.1271] [Medline: 15161774]
204. Sato E, Ochiai R, Shibayama T, Nishigaki M, Abe Y, Sawa T, et al. Reliability and validity of revised and short form versions of diabetes diet-related quality of life scale. Diabetol Int 2017 Jun;8(2):181-192 [FREE Full text] [doi: 10.1007/s13340-016-0291-x] [Medline: 30603320]
205. Hayes RP, DeLozier AM. Reliability, validity, and responsiveness of the Impact of Weight on Self-Perceptions Questionnaire (IW-SP) in individuals with type 2 diabetes and obesity. Diabetes Technol Ther 2015 Mar;17(3):210-214. [doi: 10.1089/dia.2014.0142] [Medline: 25353705]
206. Martín Payo R, González Méndez X, Papín Cano C, Suárez Álvarez J. Development and validation of a questionnaire for assessing the characteristics of diet and physical activity in patients with type 2 diabetes. Psicothema 2018 Feb;30(1):116-122. [doi: 10.7334/psicothema2017.269] [Medline: 29363480]
207. Rohani H, Eslami AA, Ghaderi A, Jafari-Koshki T, Sadeghi E, Bidkhori M, et al. Validation and psychometric evaluation of physical activity belief scale among patients with type 2 diabetes mellitus: an application of health action process approach. Health Promot Perspect 2016 Jun 11;6(2):71-79 [FREE Full text] [doi: 10.15171/hpp.2016.13] [Medline: 27386421]
208. Teng Y, Wang S, Wang N, Muhuyati. STOP-Bang questionnaire screening for obstructive sleep apnea among Chinese patients with type 2 diabetes mellitus. Arch Med Sci 2018 Aug;14(5):971-978 [FREE Full text] [doi: 10.5114/aoms.2018.73984] [Medline: 30154877]
209. Donovan LM, Yu L, Bertisch SM, Buysse DJ, Rueschman M, Patel SR. Responsiveness of patient-reported outcomes to treatment among patients with type 2 diabetes mellitus and OSA. Chest 2020 Mar;157(3):665-672 [FREE Full text] [doi: 10.1016/j.chest.2019.11.011] [Medline: 31785255]
210. Shakibazadeh E, Rashidian A, Larijani B, Shojaeezadeh D. Psychometric properties of the Iranian version of Resources and Support for Chronic Illness Self-management Scale in patients with type 2 diabetes. Int J Prev Med 2012 Feb;3(2):84-90 [FREE Full text] [Medline: 22347603]
211. Hara Y, Iwashita S, Ishii K, Inada C, Okada A, Tajiri Y, et al. The reliability and validity of the Japanese version of the Diabetes Family Behavior Checklist (DFBC) for assessing the relationship between type 2 diabetes mellitus patients and their families with respect to adherence to treatment regimen. Diabetes Res Clin Pract 2013 Jan;99(1):39-47. [doi: 10.1016/j.diabres.2012.10.014] [Medline: 23107110]
212. Littlewood K, Cummings DM, Lutes L, Solar C. Psychometric properties of the Family Support Scale adapted for African American women with type 2 diabetes mellitus. Ethn Dis 2015;25(2):193-199. [Medline: 26118148]
213. Matin H, Nadrian H, Dastjerdi L, Sarbakhsh P, Shaghaghi A. Psychometric properties of the Persian Health Care Climate Questionnaire (HCCQ-P): assessment of type 2 diabetes care supportiveness in Iran. PPA 2019 May;13:783-793. [doi: 10.2147/ppa.s201400]
214. Sofulu F, Unsalavdal RE, Arkan RB. Validity and reliability of the diabetes family support and conflict scale in Turkish. Acta Medica Mediterranea 2017;33:107-114. [doi: 10.19193/0393-6384_2017_1_017]
215. Kato A, Takada M, Hashimoto H. Reliability and validity of the Japanese version of the self-stigma scale in patients with type 2 diabetes. Health Qual Life Outcomes 2014 Dec 12;12(1):179 [FREE Full text] [doi: 10.1186/s12955-014-0179-z] [Medline: 25495723]
216. Koike M, Inagaki M, Tasaki K, Matsui K, Horiguchi T, Oda A, et al. Validation of the relationship consciousness of Japanese Patients with type 2 diabetes scale. Int J Nurs Sci 2019 Jan 10;6(1):31-37 [FREE Full text] [doi: 10.1016/j.ijnss.2018.09.007] [Medline: 31406866]
217. Hara Y, Koyama S, Morinaga T, Ito H, Kohno S, Hirai H, et al. The reliability and validity of the Japanese version of the Appraisal of Diabetes Scale for type 2 diabetes patients. Diabetes Res Clin Pract 2011 Jan;91(1):40-46. [doi: 10.1016/j.diabres.2010.09.034] [Medline: 21040993]
218. Browne JL, Ventura AD, Mosely K, Speight J. Measuring the stigma surrounding type 2 diabetes: development and validation of the type 2 Diabetes Stigma Assessment Scale (DSAS-2). Diabetes Care 2016 Dec 11;39(12):2141-2148. [doi: 10.2337/dc16-0117] [Medline: 27515964]

219. Lee E, Lee YW, Lee K, Nam M, Kim YS, Han SJ. A Korean version of the Appraisal of Diabetes Scale (ADS-K): psychometric evaluation with a population of Koreans with type 2 diabetes. J Transcult Nurs 2015 May 29;26(3):270-278. [doi: 10.1177/1043659614524793] [Medline: 24782147]
220. Ebrahimi H, Karimi Moonaghi H, Asghari Jafarabadi M, Namdar Areshtanab H, Jouybari L. Development and preliminary validation of Diabetes Adjustment Assessment Scale (DAAS): a new measure of adjustment with type 2 diabetes. J Caring Sci 2016 Jun 01;5(2):145-152 [FREE Full text] [doi: 10.15171/jcs.2016.015] [Medline: 27354978]
221. Mahjouri MY, Arzaghi SM, Qorbani M, Esfahani EN, Larijani B. Evaluation of psychometric properties of the third version of the Iranian Diabetes Attitude Scale (IR-DAS-3). J Diabetes Metab Disord 2011;10:8 [FREE Full text]
222. Persson L, Erichsen M, Wändell P, Gåfvels C. Psychometric evaluation of a coping questionnaire in two independent samples of people with diabetes. Stress Health 2013 Oct 05;29(4):286-296. [doi: 10.1002/smi.2466] [Medline: 23125028]
223. Traina S, Mathias S, Colwell H, Crosby R, Abraham C. The Diabetes Intention, Attitude, and Behavior Questionnaire: evaluation of a brief questionnaire to measure physical activity, dietary control, maintenance of a healthy weight, and psychological antecedents. Patient Prefer Adherence 2016 Feb:213. [doi: 10.2147/ppa.s94878]
224. Martinez NC, Sousa VD. Cross-cultural validation and psychometric evaluation of the Spanish Brief Religious Coping Scale (S-BRCS). J Transcult Nurs 2011 Jul 19;22(3):248-256. [doi: 10.1177/1043659611404426] [Medline: 21505097]
225. Golin CE, DiMatteo MR, Leake B, Duan N, Gelberg L. A diabetes-specific measure of patient desire to participate in medical decision making. Diabetes Educ 2001 Sep 04;27(6):875-886. [doi: 10.1177/014572170102700613] [Medline: 12211927]
226. Hearnshaw H, Wright K, Dale J, Sturt J, Vermeire E, van Royen P. Development and validation of the Diabetes Obstacles Questionnaire (DOQ) to assess obstacles in living with Type 2 diabetes. Diabet Med 2007;24:878-882. [doi: 10.1111/j.1464-5491.2007.02137.x]
227. Vandekerckhove M, Vermeire E, Weeren A, Van Royen P. Validation of the Diabetes Obstacles Questionnaire (DOQ) to assess obstacles in living with type 2 diabetes in a Belgian population. Prim Care Diabetes 2009 Feb;3(1):43-47. [doi: 10.1016/j.pcd.2009.02.003] [Medline: 19264569]
228. Pilv L, Vermeire E, Rätsep A, Moreau A, Nikolić D, Petek D, et al. Development and validation of the short version of the diabetes obstacles questionnaire (DOQ-30) in six European countries. Eur J Gen Pract 2016 Nov 18;22(1):16-22. [doi: 10.3109/13814788.2015.1093619] [Medline: 26578192]
229. Hill-Briggs F, Yeh H, Gary TL, Batts-Turner M, D'Zurilla T, Brancati FL. Diabetes problem-solving scale development in an adult, African American sample. Diabetes Educ 2007 Nov 19;33(2):291-299. [doi: 10.1177/0145721707299267] [Medline: 17426304]
230. Egede LE, Ellis C. Development and psychometric properties of the 12-item diabetes fatalism scale. J Gen Intern Med 2010 Jan 12;25(1):61-66 [FREE Full text] [doi: 10.1007/s11606-009-1168-5] [Medline: 19908102]
231. Abubakari A, Jones MC, Lauder W, Kirk A, Devendra D, Anderson J. Psychometric properties of the Revised Illness Perception Questionnaire: factor structure and reliability among African-origin populations with type 2 diabetes. Int J Nurs Stud 2012 Jun;49(6):672-681. [doi: 10.1016/j.ijnurstu.2011.11.008] [Medline: 22142936]
232. Traina SB, Colwell HH, Crosby RD, Mathias SD. Pragmatic measurement of health satisfaction in people with type 2 diabetes mellitus using the Current Health Satisfaction Questionnaire. Patientt Relat Outcome Meas 2015 Mar:103. [doi: 10.2147/prom.s79368]
233. Chew B, Vos RC, Heijmans M, Shariff-Ghazali S, Fernandez A, Rutten GEHM. Validity and reliability of a Malay version of the brief illness perception questionnaire for patients with type 2 diabetes mellitus. BMC Med Res Methodol 2017 Aug 03;17(1):118 [FREE Full text] [doi: 10.1186/s12874-017-0394-5] [Medline: 28774271]
234. 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 healthcare interventions: explanation and elaboration. BMJ 2009 Jul 21;339(jul21 1):b2700-b2700 [FREE Full text] [doi: 10.1136/bmj.b2700] [Medline: 19622552]
235. Terwee CB, Prinsen CAC, Ricci Garotti MG, Suman A, de Vet HCW, Mokkink LB. The quality of systematic reviews of health-related outcome measurement instruments. Qual Life Res 2016 Apr 7;25(4):767-779 [FREE Full text] [doi: 10.1007/s11136-015-1122-4] [Medline: 26346986]

Abbreviations

ADDQoL-19: 19-item Audit of Diabetes-Dependent Quality of Life
CH-ASIQ: Chinese Attitudes to Starting Insulin Questionnaire
CIRS: Chronic Illness Resources Survey
C-ITAS-HK: Insulin Treatment Appraisal Scale in Hong Kong primary care patients
COSMIN: COnsensus-based Standards for the selection of health Measurement Instruments
C-PDQ: Chinese version of the Personal Diabetes Questionnaire
CUDOS-Chinese: Mandarin Chinese Version of the Clinically Useful Depression Outcome Scale
DAAS: Diabetes Adjustment Assessment Scale
DES-SF: Diabetes Empowerment Scale–Short Form

DHPSC: diabetes health promotion self-care scale
DMKT: Diabetes Mellitus Knowledge Test
DMSES: diabetes management self-efficacy scale
DOQ: Diabetes Obstacles Questionnaire
DPMD: diabetes-specific measure of patient desire to participate in medical decision making
DQOL: Diabetes Quality-of-Life Measure
DRNK: diabetes-related nutrition knowledge questionnaire
DSMI-20: Diabetes Self-Management Instrument Short Form
DSMQ: Diabetes Self-management Questionnaire
DTSQ: Diabetes Treatment Satisfaction Questionnaire
ED-5Q: EuroQol 5-Dimension
FACIT: Functional Assessment of Chronic Illness Therapy
FH-15: new fear of hypoglycemia scale
GRADE: Grading of Recommendations Assessment, Development and Evaluation
HAPA: health action process approach
HCCQ-P: Persian Health Care Climate Questionnaire
HSQ 2.0: Health Status Questionnaire 2.0
KHLS-DM: Korean Health Literacy Scale for Diabetes Mellitus
MAT insulin: Measurement of Adherence to Drug Therapy in Diabetes Mellitus–Insulin Therapy
MAT OADS: Measurement of Adherence to Drug Therapy in Diabetes Mellitus–Oral Antidiabetics
PACIC: Patient Assessment of Chronic Illness Care
PAID: Problem Areas in Diabetes scale
PAID-5: short form of the Problem Areas in Diabetes scale
PCSD-P: Persian Version of Perceived Competence Scale for Diabetes
PHQ-9: Patient Health Questionnaire-9
PRISMA: Preferred Reporting Items for Systematic Review and Meta-Analysis
PRO-DM-Thai: instrument for patient-reported outcomes in Thai patients with type 2 diabetes mellitus
PROM: patient-reported outcome measure
T2DHPS: Type 2 Diabetes and Health Promotion Scale
WHOQOL-100: World Health Organization Quality of Life questionnaire

Edited by R Kukafka; submitted 14.10.20; peer-reviewed by H Goh, Z Zrubka; comments to author 26.11.20; revised version received 18.12.20; accepted 14.06.21; published 13.08.21.

Please cite as:
Wee PJL, Kwan YH, Loh DHF, Phang JK, Puar TH, Østbye T, Thumboo J, Yoon S, Low LL
Measurement Properties of Patient-Reported Outcome Measures for Diabetes: Systematic Review
J Med Internet Res 2021;23(8):e25002
URL: https://www.jmir.org/2021/8/e25002
doi:10.2196/25002
PMID:34397387

©Priscilla Jia Ling Wee, Yu Heng Kwan, Dionne Hui Fang Loh, Jie Kie Phang, Troy H Puar, Truls Østbye, Julian Thumboo, Sungwon Yoon, Lian Leng Low. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Patient and Clinician Characteristics Associated With Secure Message Content: Retrospective Cohort Study

Dawn Heisey-Grove[1], MPH, PhD; Cheryl Rathert[2], PhD; Laura E McClelland[3], PhD; Kevin Jackson[4], RRT, PhD; Jonathan P DeShazo[3], MPH, PhD

[1]The MITRE Corporation, McLean, VA, United States

[2]Department of Health Management and Policy, College for Public Health and Social Justice, Saint Louis University, St. Louis, MO, United States

[3]Department of Health Administration, College of Health Professions, Virginia Commonwealth University, Richmond, VA, United States

[4]Health Services Management, Department of Nursing and Allied Health, Norfolk State University, Norfolk, VA, United States

Corresponding Author:
Dawn Heisey-Grove, MPH, PhD
The MITRE Corporation
7525 Colshire Dr
McLean, VA, 22102
United States
Phone: 1 7035477389
Email: heiseygroved@mitre.org

Abstract

Background: Good communication has been shown to affect patient outcomes; however, the effect varies according to patient and clinician characteristics. To date, no research has explored the differences in the content of secure messages based on these characteristics.

Objective: This study aims to explore characteristics of patients and clinic staff associated with the content exchanged in secure messages.

Methods: We coded 18,309 messages that were part of threads initiated by 1031 patients with hypertension, diabetes, or both conditions, in communication with 711 staff members. We conducted four sets of analyses to identify associations between patient characteristics and the types of messages they sent, staff characteristics and the types of messages they sent, staff characteristics and the types of messages patients sent to them, and patient characteristics and the types of messages they received from staff. Logistic regression was used to estimate the strength of the associations.

Results: We found that younger patients had reduced odds of sharing clinical updates (odds ratio [OR] 0.77, 95% CI 0.65-0.91) and requesting prescription refills (OR 0.77, 95% CI 0.65-0.90). Women had reduced odds of self-reporting biometrics (OR 0.78, 95% CI 0.62-0.98) but greater odds of responding to a clinician (OR 1.20, 95% CI 1.02-1.42) and seeking medical guidance (OR 1.19, 95% CI 1.01-1.40). Compared with White patients, Black patients had greater odds of requesting preventive care (OR 2.68, 95% CI 1.30-5.51) but reduced odds of requesting a new or changed prescription (OR 0.72, 95% CI 0.53-0.98) or laboratory or other diagnostic procedures (OR 0.66, 95% CI 0.46-0.95). Staff had lower odds of sharing medical guidance with younger patients (OR 0.83, 95% CI 0.69-1.00) and uninsured patients (OR 0.21, 95% CI 0.06-0.73) but had greater odds of sharing medical guidance with patients with public payers (OR 2.03, 95% CI 1.26-3.25) compared with patients with private payers. Staff had reduced odds of confirming to women that their requests were fulfilled (OR 0.82, 95% CI 0.69-0.98). Compared with physicians, nurse practitioners had greater odds of sharing medical guidance with patients (OR 2.74, 95% CI 1.12-6.68) and receiving prescription refill requests (OR 3.39, 95% CI 1.49-7.71). Registered nurses had greater odds of deferred information sharing (OR 1.61, 95% CI 1.04-2.49) and receiving responses to messages (OR 3.93, 95% CI 2.18-7.11) than physicians.

Conclusions: The differences we found in content use based on patient characteristics could lead to the exacerbation of health disparities when content is associated with health outcomes. Disparities in the content of secure messages could exacerbate disparities in patient outcomes, such as satisfaction, trust in the system, self-care, and health outcomes. Staff and administrators should evaluate how secure messaging is used to ensure that disparities in care are not perpetuated via this communication modality.

(J Med Internet Res 2021;23(8):e26650) doi:10.2196/26650

KEYWORDS
patient-provider communication; electronic messaging; hypertension; diabetes

Introduction

Background

Appropriate use of health information technology may promote patient engagement and empowerment by improving patients' preparation for, and recall of, clinical encounters [1]. One form of health information technology is secure messaging—the electronic exchange of messages between patients and clinicians, typically via a secure platform such as a patient portal. Published research highlights the potential of secure messaging to support patient satisfaction, access to care, and health outcomes. Most research has explored health care utilization, with a number of studies identifying reductions in patients' visits associated with secure messaging [2-4]. Other studies have identified improvements in selected measures for screening and testing associated with secure messaging use [5-7]. Secure message use has also been associated with improvements in blood pressure control [5,8,9], glycemic levels [5-7,10-12], and improved postdischarge coping [13].

However, secure messaging use is associated with a variety of clinician and patient characteristics [14-16]. Furthermore, moderators between communication and patient health outcomes include both patient and clinician characteristics (eg, age, gender, race, income, and education) [17]. Consistent with this, research indicates that differential use by patients' race and ethnicity persists once patients access patient portals [15]. If secure messaging is associated with improvements in patients' satisfaction, access to care, and health outcomes but use varies according to patient and clinician characteristics, there is a chance that the benefits of secure messaging communication may be inequitably applied across populations, leaving some patients without the benefits of that form of communication.

Communication functions such as information exchange, emotional support, uncertainty management, and support for decision-making and self-management can be provided by clinicians through secure messaging, leading to changes in patients' health outcomes [18]. Research has demonstrated associations between patients' improved glycemic levels and diastolic blood pressure and clinicians' information-sharing message content [12]. The same study found that negative message content (eg, denying patients' requests and responses deferring answers to a later time) was associated with increased systolic blood pressure. Although published work demonstrates that communication strategies vary according to age, sex, race, primary language, and comfort level with the communication medium [19-23], the authors could find no research on whether differences in message content exist based on patient or clinician characteristics.

Objectives

In this study, we explore whether differences exist in communication functions based on characteristics of patients and clinicians, representing the senders and receivers in secure message threads. Using a taxonomy created specifically for secure messages, we coded the patient- and staff-generated messages in a large sample of patient-initiated message threads [24]. We then explored the differences in message content prevalence based on the characteristics of senders and receivers. Our hypotheses for this research are as follows:

- Hypothesis 1 (patients as senders): message content sent by patients to staff will vary based on patients' age, sex, race, health status, insurance type, and proximity to the clinic.
- Hypothesis 2 (staff as receivers): patients will vary their message content based on the staff type and clinical specialty of the intended recipient.
- Hypothesis 3 (staff as senders): message content in staffs' replies will vary based on staff type, clinical specialty, and annual message volume.
- Hypothesis 4 (patients as receivers): staff will vary their message content based on patients' age, sex, race, health status, and insurance type.

Methods

Study Population

Our study included adult patients with diabetes, hypertension, or both conditions selected from patients of a large urban medical center who sent secure messages using the outpatient portal of the medical center (Cerner) between January 1 and December 31, 2017. To ensure that patients were patients at the medical center for the study duration and their diagnoses persisted during that period, we included patients with relevant diagnosis codes in the years preceding (2016) and following (2018) the study period. Patients had to have at least two outpatient visits or 1 inpatient visit in 2016 with diagnosis codes for either diabetes (*International Classification of Diseases, Tenth Revision, Clinical Modification* [*ICD-10-CM*] E11) or hypertension (*ICD-10-CM* I10), and at least one outpatient visit between January and June 2018. We only included visits within the medical center.

We stratified patients who met the inclusion criteria based on their health condition (hypertension only, diabetes only, or both conditions). The required sample size necessitated the inclusion of all patients with diabetes. We used version 9.2 of the SAS System for Windows (SAS Institute Inc) to select a simple random sample from each of the other two strata. We then included all staff who were the intended recipients of, or who responded to, our sampled patients' secure messages during the study period.

Our analyses included all threads initiated by the sampled patients, completed, and saved to patients' charts between January 1 and December 31, 2017. Secure messages were extracted during a chart review of each patient's electronic medical record. We did not include communications outside secure messaging in these analyses. This research was approved by the institutional review board.

Table S1 of Multimedia Appendix 1 includes the number and percentage of patients sampled and the census counts of messages and staff senders and receivers. Our patient study population included 1031 patients who generated 7346 patient-initiated threads during 2017. Our staff population was 711; of those, 56.6% (403/711) sent and were the intended recipients of at least one message. Our message sample included 18,309 messages, of which slightly more than half (10,163/18,309, 55.55%) were patient-generated.

Patient Characteristics

We included categorical variables representing patients' demographic and geography-based characteristics and elements for health status and health care access. Demographic characteristics included age, sex, and race (Black, White, and other). Geography-based characteristics included rural or urban home locations based on rural-urban commuting area [25] codes and average travel distance in miles between clinic and home.

We included patients' health status markers based on health conditions (ie, diabetes, hypertension, or both conditions) and the number of comorbidities ranging from 1 to 9 from a list of *ICD-10-CM* that frequently occurred within the sampled population. Finally, we incorporated proxy elements for patients' health care access using payer type (private, public, uninsured, or other) and the number of outpatient visits in 2017. All analyses controlled for the number of threads initiated by patients in 2017 because the greater the number of messages sent by a patient, the greater the opportunity for a variety of taxa.

Characteristics of Staff

Clinic teams at the medical center typically triage patient messages; thus, the intended recipient was not always the individual who responded to a given message. Our analyses, therefore, differentially identify for each patient-generated message the staff who sent the message response from the receiver as the intended recipient of patient-generated message.

We used two strategies to identify staff receivers as staff to whom the patient intended the message to be delivered. First, for the initial patient-generated message in each thread, we identified the receiver as the staff to whom the message was addressed. Second, we assumed that the receiver for all subsequent patient-generated messages was the sender of the staff-generated message that most recently preceded the patient-generated message. If a staff-generated message did not precede the patient-generated message, we used the same receiver as the most recently preceding patient-generated message.

We included three variables to classify staff. First, staff types were grouped into the 6 most frequently occurring types (ie, administrative staff, licensed practical nurses, nurse practitioners, physicians, registered nurses, and other clinicians). The *other* category included pharmacists, physician assistants, medical assistants, podiatrists, social workers, and medical technicians. Next, we categorized clinical specialty as either primary care or specialty. We included family and internal medicine, geriatrics, pediatrics, obstetrics, and gynecology in our primary care category. Physician assistants, registered nurses, pharmacists, social workers, medical technicians, case managers, counselors, and administrative staff were not assigned a specialty.

Finally, we estimated the message volume for each staff member based on the messages saved to all patients' charts (not just our sampled population), regardless of whether they were sent in response to a patient-initiated thread or were part of a staff-initiated thread.

Content Analysis

Consistent with the premises of the Uncertainty in Illness Theory [26] and patient-centered communication [18], our taxonomy includes codes (or taxa) for patients seeking information to alleviate uncertainty around their health status (eg, symptoms and condition) and health care delivery processes. It also includes task-oriented requests that may be used to support self-care or address uncertainty. We included social communication and information-sharing taxa for both patient- and staff-generated messages because these taxa may indicate communication that fosters trust-building between patients and clinicians. For content from staff, the taxonomy also includes action responses based on the taxonomy of requests by patients [27] as leveraged by other researchers. Additional taxa for staff-generated messages classified clinicians' information-sharing content.

More details on the content analysis process are provided elsewhere [24]; however, in summary, a primary coder read and assigned taxa to all messages, and a second coder did the same for a random 10% sample of messages. Coding units could be no longer than a single message and were frequently shorter, with multiple codes applied to a single message. Each taxon was assigned only once to a given message. We coded the data using NVivo 12 software (QSR International). Discrepancies were reconciled, and the primary coder recoded the messages accordingly.

Data Analysis

We explored the associations between taxa and characteristics of senders and receivers. For each taxon (ie, individual code), we created a set of dichotomous variables: one set based on the sender and the other on the receiver. For sender-based analyses, we recorded the variable as positive if the patient or staff sent at least one patient-generated or staff-generated message coded with the taxon, respectively. For receiver-based analyses, we assigned a positive value if the patient or staff received at least one staff-generated or patient-generated message coded with the taxon, respectively.

We estimated adjusted odds ratios (ORs) using separate logistic regression models, where each taxon was the dependent variable, and the patient or staff characteristics were the independent variables. Analyses were conducted using version 9.2 of the SAS System for Windows.

Results

Population Characteristics

Table 1 shows patient characteristics based on their health conditions. On average, patients sent 9.86 messages (SD 13.70;

median 1.0; maximum 117) across 7.12 (SD 9.66) threads. Our population primarily lived in urban areas and comprised approximately two-thirds of women.

Table 2 presents the characteristics of the staff as both receivers and senders. Patients directed more messages to physicians and primary care clinicians. Registered nurses were the most common type of sender staff, followed by physicians. Although 3.69% (376/10,163) of the messages were addressed to administrative staff, those staff accounted for almost a quarter of the messages sent.

Table 1. Patient characteristics by health condition (N=1031).

Characteristics	Diabetes only (n=398)	Hypertension only (n=394)	Both conditions (n=239)	Total (N=1031)
Number of messages, mean (SD)	9.93 (13.57)	9.12 (12.91)	10.95 (15.11)	9.86 (13.70)
Number of threads, mean (SD)	7.07 (8.96)	6.54 (9.07)	8.18 (11.51)	7.12 (9.66)
Age (years), mean (SD)	54.65 (13.83)	59.62 (14.34)	60.49 (12.01)	57.91 (13.87)
Distance between home and clinic (miles), mean (SD)	26.88 (38.36)	34.83 (37.62)	31.62 (41.88)	31.02 (39.05)
Number of co-occurring conditions, mean (SD)	3.02 (1.83)	2.95 (1.67)	4.32 (1.88)	3.29 (1.87)
Number of outpatient visits, mean (SD)	14.31 (12.04)	15.54 (12.23)	19.15 (14.29)	15.90 (12.79)
Female, n (%)	277 (69.6)	239 (60.7)	154 (64.4)	670 (65)
Urban home location, n (%)	392 (98.5)	380 (96.4)	233 (97.5)	1005 (97.5)
Insurance, n (%)				
Other	109 (27.4)	87 (22.1)	75 (31.4)	271 (26.3)
Private	174 (43.7)	113 (28.7)	44 (18.4)	331 (32.1)
Public	108 (27.1)	189 (48)	115 (48.1)	412 (40)
Uninsured	7 (1.8)	5 (1.3)	5 (2.1)	17 (1.6)
Race, n (%)				
Black	182 (45.7)	130 (33.2)	104 (43.5)	416 (40.4)
Other	26 (6.5)	12 (3.1)	12 (5)	50 (4.9)
White	190 (47.7)	250 (63.8)	123 (51.5)	563 (54.7)

Table 2. Staff characteristics (N=711).

Characteristics	Receivers, n (%)		Senders, n (%)		Total staff (N=711), n (%)
	Staff (n=567)	Messages (n=10,163)	Staff (n=544)	Messages (n=8146)	
Staff type					
Administrative	40 (7.1)	376 (3.7)	79 (14.5)	1927 (23.7)	79 (11.1)
Licensed practical nurse	17 (3)	148 (1.5)	32 (5.9)	474 (5.8)	32 (4.5)
Nurse practitioner	63 (11.1)	918 (9)	50 (9.2)	503 (6.2)	65 (9.1)
Other staff type	27 (4.8)	136 (1.3)	32 (5.9)	158 (1.9)	37 (5.2)
Registered nurse	114 (20.1)	1222 (12)	169 (31)	2678 (32.9)	170 (23.9)
Physician	294 (51.9)	5736 (56.4)	163 (30)	2380 (29.2)	304 (42.8)
Unknown	12 (2.1)	1627 (16)	19 (3.5)	26 (0.3)	24 (3.4)
Clinical specialty					
N/A[a]	224 (39.5)	1975 (19.4)	324 (59.6)	5176 (63.5)	346 (48.7)
Primary	155 (27.3)	4387 (43.2)	116 (21.3)	2002 (24.6)	159 (22.4)
Specialty	174 (30.7)	2159 (21.2)	81 (14.9)	937 (11.5)	178 (25)
Unknown	14 (2.5)	1642 (16.2)	23 (4.2)	31 (0.4)	28 (3.9)
Messages in 2017					
≤1000	206 (36.3)	1092 (10.7)	153 (28.1)	479 (5.9)	267 (37.6)
1001-2000	134 (23.6)	1320 (13)	134 (24.6)	789 (9.7)	166 (23.3)
2001-3400	99 (17.5)	1859 (18.3)	109 (20)	1244 (15.3)	119 (16.7)
>3400	109 (19.2)	4240 (41.7)	121 (22.2)	5550 (68.1)	123 (17.3)
Unknown	19 (3.4)	1652 (16.3)	27 (5)	84 (1)	36 (5.1)

[a] N/A: not applicable.

Patient-Generated Taxa

Overview

Table 3 presents a description of each patient-generated taxon, the proportion of patients who sent them, and staff who received these taxa. Most patients sent messages with *information sharing* and *information seeking* content. The lowest percentage of patients sent *social communication* (*appreciation* or *praise*) and preventive care scheduling requests. Furthermore, 3 of 4 staff received at least one *information sharing* and one *information seeking* message. The smallest percentage of staff received appointment cancellation requests.

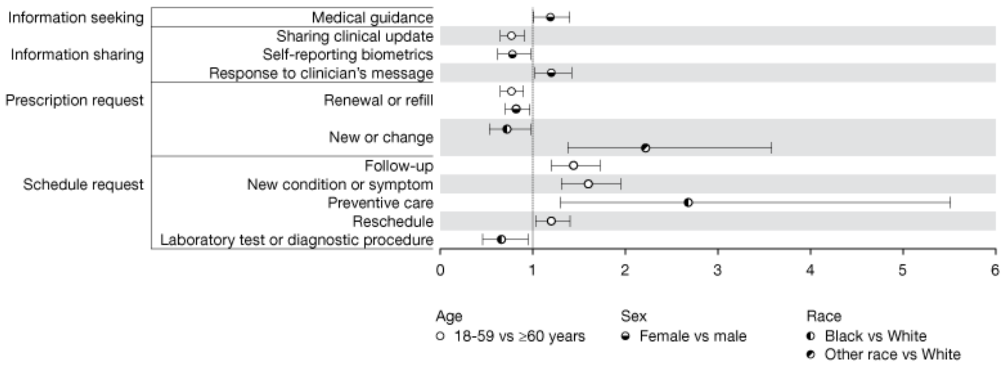
Table 3. Percentage of patient-generated taxa by patient senders and staff recipients.

Patient-generated taxon	Description	Patient senders (N=1031), n (%)	Intended staff recipients (n=567), n (%)
Information seeking			
Logistics	Questions about timing, clinical processes, health care settings, or a patient's care plan	420 (40.7)	306 (54.0)
Medical guidance	Questions that seek medical guidance or information	554 (53.7)	329 (58.0)
Information sharing			
Sharing clinical update	Sharing information with a clinician that does not require immediate action or response (and may not require action at all)	479 (46.4)	297 (52.4)
Self-reporting	Sharing biometrics or other health-related self-measurements; information with a clinician that does not require immediate action or a response	109 (10.6)	66 (11.6)
Response to clinician's message	Response to clinician's question in a preceding message within the thread	520 (50.4)	258 (45.5)
Prescription request			
Prescription refill or renewal	Request for prescription refill or renewal	504 (48.9)	246 (43.4)
New or change prescription	Request for a new prescription or switch to a different medication or treatment	293 (28.4)	164 (28.9)
Referral request	Request for a referral to another health care facility or clinician	116 (11.2)	94 (16.6)
Other administrative requests	Process-related and administrative in nature; includes requests for sick notes, contact information, medical records, patient portal access, or information about billing or insurance; technology-related questions; requests for call	312 (30.3)	230 (40.6)
Scheduling request			
Cancellation	Request to cancel an existing appointment with no associated request to change the date or time	213 (20.6)	39 (6.9)
Follow-up	Request for an appointment relative to an existing health condition	241 (23.4)	92 (16.2)
New condition or symptom	Patient request for an appointment relative to a newly identified health condition or new symptom for an existing condition; new patient appointment	200 (19.4)	76 (13.4)
Preventive care	Request for preventive care or routine physical examination	85 (8.2)	17 (3.0)
Reschedule	Request for an appointment to be changed to another date or time	393 (38.1)	112 (19.8)
Laboratory test or diagnostic procedure	Request for a laboratory test or diagnostic procedure (eg, x-ray or ultrasound) order	145 (14.1)	104 (18.3)
Social communication			
Appreciation or praise	Content that expresses gratitude or offers acknowledgment or appreciation of service provided, health status, or another act	65 (6.3)	56 (9.8)
Complaints	Expressions of frustration or displeasure about service or life issues	96 (9.3)	107 (18.9)
Life issues	Communication about aspects of the patients' life not specifically related to health	125 (12.1)	94 (16.6)

Characteristics of Patients as Senders

Figure 1 displays the ORs estimated as statistically significant with a *P*<.05 for the associations between patient-generated taxa and patient demographic characteristics. Multimedia Appendix 1 presents the OR estimates for all taxa.

Figure 1. Odds ratios and 95% CIs of associations between patient demographic characteristics and patient-generated message content.

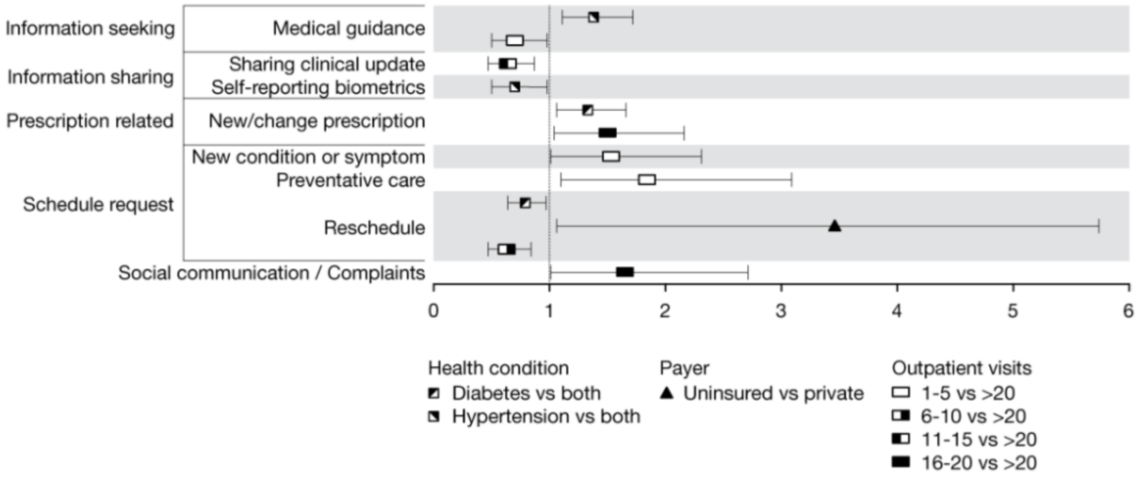


Younger patients had lower odds of sending clinical updates (OR 0.77, 95% CI 0.65-0.91) and prescription refill and renewal requests (OR 0.77, 95% CI 0.65-0.90). They had higher odds of sending scheduling requests, specifically for follow-up appointments (OR 1.44, 95% CI 1.20-1.73), appointments for new conditions or symptoms (OR 1.60, 95% CI 1.31-1.95), and rescheduling (OR 1.20, 95% CI 1.03-1.41). Women had lower odds of self-reporting biometrics (OR 0.78, 95% CI 0.62-0.98) and requesting prescription refills (OR 0.82, 95% CI 0.70-0.97), but higher odds of responding to staffs' comments or questions (OR 1.20, 95% CI 1.02-1.42) and seeking medical guidance (OR 1.19, 95% CI 1.01-1.40).

Black patients had lower odds of requesting a new or changed medication (OR 0.72, 95% CI 0.53-0.98), scheduling a laboratory or other diagnostic procedure (OR 0.66, 95% CI 0.46-0.95), and requesting an appointment be canceled (OR 0.73, 95% CI 0.53-1.00) compared with White patients. Conversely, Black patients had greater odds (OR 2.68, 95% CI 1.30-5.51) of requesting preventive care appointments than White patients. Patients of other races had greater odds (OR 2.2, 95% CI 1.38-3.58) of requesting a new or changed medication compared with White patients.

Figure 2 presents the ORs for patients' health care access and health status characteristics. Uninsured patients had greater odds (OR 2.46, 95% CI 1.06-5.74) of requesting an appointment to be rescheduled than patients with private payers. Patients with diabetes only had greater odds of requesting a new or changed medication (OR 1.33, 95% CI 1.06-1.66) and reduced odds of requesting an appointment to be rescheduled (OR 0.79, 95% CI 0.64-0.97) compared with patients with both diabetes and hypertension. Patients with hypertension only had greater odds of seeking medical guidance (OR 1.38, 95% CI 1.11-1.72) and reduced odds of self-reporting biometrics (OR 0.70, 95% CI 0.50-0.98) than patients with both conditions.

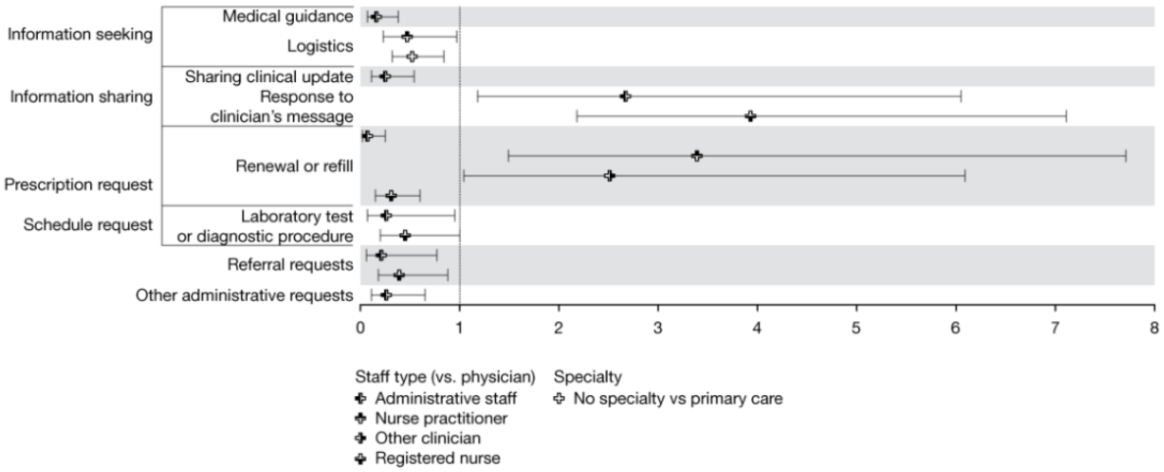
Figure 2. Odds ratios and 95% CIs of associations between patient-generated message content and patient health condition and delivery characteristics.



Characteristics of Staff as Receivers

An average of 4.92 (SD 7.59) sampled patients sent each staff 15.81 (SD 36.93) messages across 12.92 (SD 27.66) threads. Figure 3 displays the OR estimates of the associations between the characteristics of the staff as receivers of patient-generated taxa. The only differences we observed by specialty were between staff with no applicable specialty and primary care clinicians. Staff not assigned a specialty had lower odds of receiving *logistics* requests (OR 0.52, 95% CI 0.32-0.84) from patients. The administrative staff were less likely to receive medical guidance requests (*P*<.001), clinical updates (*P*<.001), prescription refill requests (*P*<.001), laboratory or other procedure scheduling requests (*P*=.04), and other administrative (*P*=.003) and referral requests (*P*=.02) than physicians. They had greater odds (OR 2.67, 95% CI 1.18-6.05) of receiving responses to their questions. Registered nurses also had greater odds of receiving *response to the clinician's message* (OR 3.93, 95% CI 2.18-7.11) and lower odds of receiving requests for referrals (*P*=.02) and refilling prescriptions (*P*<.001). Nurse practitioners had greater odds (OR 3.39, 95% CI 1.49-7.71) of receiving prescription-related requests but had lower odds of receiving *logistics* questions (OR 0.47, 95% CI 0.23-0.97) than physicians.

Figure 3. Odds ratios and 95% CIs of associations between clinical staff characteristics and patient-generated content.



Staff-Generated Taxa

Overview

Table 4 lists the percentages of staff who sent at least one message with the selected taxon. Almost 9 in 10 staff (n=473) shared information with their patients, although only slightly more than half shared medical guidance. More than half of staff sent at least one message that fulfilled a patient's request.

Three-quarters of the patients (n=798) received at least one message from the staff with information-sharing content. Two-thirds of the patients received message content that fulfilled their request. Few patients received messages that denied their requests or provided encouragement.

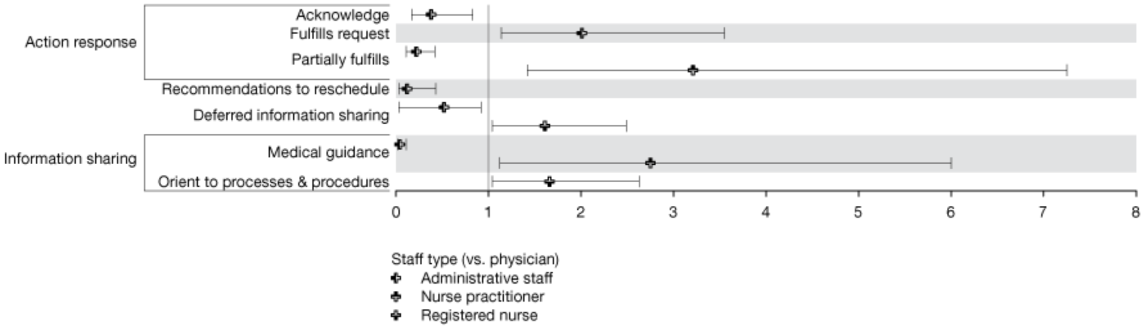
Table 4. Staff-generated taxa distribution.

Staff-generated taxon	Description	Staff senders (n=544), n (%)	Patient recipients (N=1031), n (%)
Action response			
Acknowledges	Includes recognition that the request for action or information is made, but no indication is provided about whether the request will be fulfilled	148 (27.2)	254 (24.6)
Fulfills request	Documentation that the request action was completed	316 (58.1)	686 (66.5)
Partially fulfills	Indicates additional steps are necessary to fulfill the request or that only part of the request can or has been completed	161 (29.6)	283 (27.4)
Denies request	Indicates that the request will not be fulfilled	57 (10.5)	95 (9.2)
Information seeking	Clinicians' requests for information or clarity around patients' condition or symptoms, or symptom severity or duration	248 (45.6)	552 (53.5)
Deferred information sharing	Clinical responses that refer the patient to another clinician for a response; postpone an answer pending additional clinical information	248 (45.6)	503 (48.8)
Information sharing			
Medical guidance	Provides treatment decisions, gives care instructions, dietary guidance, instructs the patient on the best next steps in their care plan, interprets diagnostic procedure or laboratory results, or provides information on symptoms or the patient's health condition	299 (55)	503 (48.8)
Orientation to procedures, treatments, or preventive behaviors	Explains what a patient might expect during treatment or diagnostic procedure or in a new health care setting or situation	371 (68.2)	718 (69.6)
Recommendation to schedule an appointment	Suggestion that patient schedule an appointment	113 (20.8)	170 (16.5)
Social communication or encouragement	Provides positive reinforcement of patient's actions or behaviors	38 (7)	58 (5.6)

Characteristics of Staff as Senders

Staff responded, on average, with 15.47 (SD 41.12) messages to 11.60 (SD 30.69) threads initiated by 6.69 (SD 17.83) patients. Figure 4 presents the estimates of the associations between taxa and staff characteristics. We observed no associations between taxa and clinical specialty after controlling for staff type and message volume. Administrative staff had reduced odds of sending many taxa compared with physicians, except for *fulfills request* (OR 2.01, 95% CI 1.14-3.55). Nurse practitioners had greater odds (OR 2.74, 95% CI 1.12-6.68) of sharing medical guidance with patients and greater odds (OR 3.21, 95% CI 1.42-7.25) of partially fulfilling patients' requests compared with physicians. Registered nurses also had greater odds of deferred information sharing (OR 1.61, 95% CI 1.04-2.49) and *orientation to procedures, treatments, and preventive behaviors* (OR 1.66, 95% CI 1.04-2.63) compared with physicians.

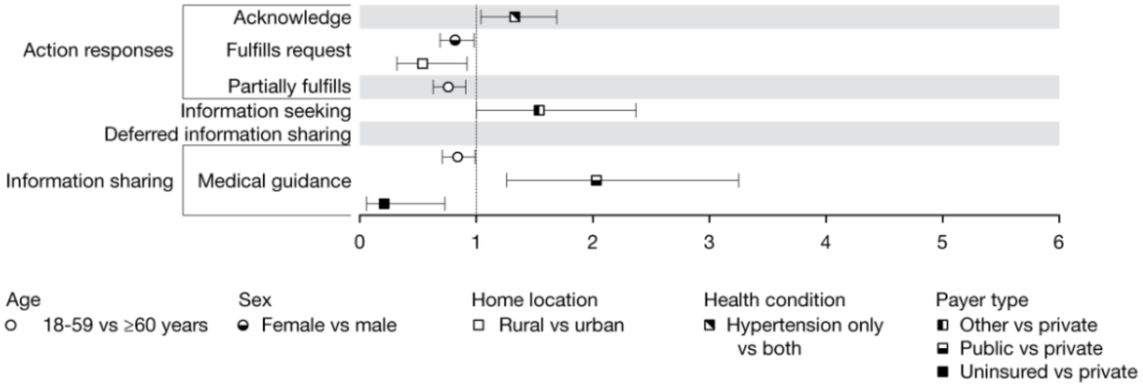
Figure 4. Odds ratios and 95% CIs of associations between clinical staff characteristics and staff-generated content.



Characteristics of Patients as Receivers

Figure 5 displays the OR estimates for associations between clinician-generated taxa and the characteristics of patients who received these taxa. Younger patients had reduced odds of receiving partial request fulfillment (OR 0.76, 95% CI 0.63-0.91) and medical guidance (OR 0.84, 95% CI 0.71-0.99). Women and individuals with a rural home address had reduced odds (OR 0.82, 95% CI 0.69-0.98 and OR 0.54, 95% CI 0.32-0.92, respectively) of receiving confirmation that their requests were fulfilled. Patients with public payers had more than twofold increased odds (OR 2.03, 95% CI 1.26-3.25), whereas uninsured patients had reduced odds (OR 0.21, 95% CI 0.06-0.73) of receiving medical guidance compared with patients with private payers.

Figure 5. Odds ratios and 95% CIs of associations between patient characteristics and staff-generated content.



Discussion

Principal Findings

Overview

As expected, secure message content varied based on the characteristics of both the sender and receiver. The patients' message content varied based on age, sex, home location, insurance type, and health condition. Staff-generated content varied primarily based on staff type. Message content staff sent to patients varied based on the patients' age, sex, health condition, and payer status. Finally, patients sent different content based on staff type. Given that other research demonstrated that secure message content was associated with selected health outcomes [28], our findings may indicate that inequitable use of secure messaging could further compound existing disparities in health care delivery and outcomes.

Patient-as-Sender

Patients who trust their clinicians may be more open to sharing information with their clinicians [17]. Previous research reported a positive association between patients' age and trust in their clinicians [29] and found that non-Hispanic White patients and men were more likely to disclose information to clinicians [30]. Consistent with these findings, we observed that younger patients were less likely to share clinical updates with their clinical team, and women were less likely than men to self-report biometrics through secure messaging. As sharing relevant clinical information with the care team can be important to the continuity of care and ongoing patient engagement, it will be important to better understand why these populations might not be taking advantage of secure messaging in this way.

Our findings indicated no difference by race for information-seeking and information-sharing content, contrary to other studies that found that Black and other race patients reported higher levels of trust in other information sources (eg, charitable organizations, newspapers, and radio) [31]. As there is an existing divide by race in the use of secure messaging [14-16], there may be less observable differences by race in the content among the patients who opt to use secure messaging. That is, patients who opt to use secure messaging are those who trust their clinic staff to provide the information they need more so than those other sources. Further research should explore the association of trust among users of secure messaging.

Our study found that Black patients were less likely to request changes to their prescriptions or request laboratory or other diagnostic procedures, whereas patients of other races were more likely to request prescription changes than White patients. These 2 request types, unlike some other task-oriented request taxa, involve a more active involvement from the patient to be aware of a medical need and outreach to the clinician to request clinical action for a change in care. Two-thirds of studies in a literature review of the effects of race on patient-physician communication reported that Black patients had fewer acts of participation during their physician visits [32]. If requests for a new or changed medication, laboratory, or other diagnostic procedures are considered more participatory in nature, then our research extends these findings to electronic communication modalities.

Staff-as-Receiver

Differences in the types of messages sent by staff were likely reflective of the fact that many practices triage messages through a team of nurses, physician assistants, pharmacists, and physicians, with physicians generally responding only to the more complicated messages [33-35]. We observed that patients were more likely to send prescription requests to nurse practitioners than physicians and were more likely to send referral and laboratory and diagnostic procedure requests to physicians, which is consistent with a triage response system.

As expected, patients intended most of their information-seeking messages to be received by physicians, nurse practitioners, and registered nurses. Although there were no differences by staff type for the staff-generated *information seeking* taxon, patients were almost four times more likely to send *responses to clinician's messages* to registered nurses and three times more likely to send them to administrative staff. This could indicate one of two factors: (1) those staff types asked more questions about patients or (2) those staff types were better at soliciting responses from patients. As noted previously, patient information sharing is a marker of trust with the clinical provider, so higher occurrences of the patient and clinical team engaging in electronic bidirectional dialog represented by this taxon might lead to strong trust or be a marker of existing trust. Alternatively, registered nurses and administrative staff sent high volumes of secure messages, so they may be more comfortable with the communication modality and better able to ask questions in a manner in which the patient is comfortable. Future studies should incorporate experience with secure messaging to control for this potential confounder.

Staff-as-Sender

Consistent with the triaging process, we found that administrative staff were less likely than physicians to share information and make recommendations to schedule appointments. In a triage system where physicians generally respond to the most complex messages, it makes sense that registered nurses and nurse practitioners were more likely than physicians to send most types of messages, as our data showed.

Previous research showed that almost 2 out of 10 office visits with a primary care physician were suitable for another modality [36]. Our research demonstrates that much of information sharing and action responses to messages is handled by registered nurses and nurse practitioners, although physicians still send the second highest number of messages. As messages could be coded with more than one taxon, it is possible that nurse respondents sent messages that addressed more than one content area, compared with physicians whose responses may have been more targeted.

Patient-as-Receiver

Younger patients were less likely to receive acknowledgment and indications for partial fulfillment. We observed differences according to age for patients' task-oriented requests (eg, scheduling, prescription-related, and administrative requests), although directionality varied (eg, younger patients were more likely to send scheduling requests but less likely to make prescription requests). It may be that the difference in action responses from the staff was associated with the preceding request type. It is unclear whether these data represent differential fulfillment rates by request type or a difference in the way staff communicate based on patient age.

Similarly, although we identified only one difference by patient sex associated with sending task-oriented requests, women were less likely to receive fulfillment responses. Further research is needed to determine if differences in fulfillment rates are based on patients' sex or the nature of the requests made by the patient. Research that explores the differences in responses among subsets of patients who sent messages with selected taxa could determine whether these responses vary among patients requesting that type of information. For example, do staff respond to prescription requests differently based on patient characteristics, whereas scheduling requests receive standard responses regardless of patient demographics? Our research did not explore the paired call-response nature of the secure message thread. Future research should explore the best approach to analyzing paired taxa in threads to understand the associations between a patient request and the staff response to that request.

The staffs' message responses did not vary by patient race. The literature on differences in patient-clinician communication by race is mixed, but a recent literature review found that most vignette studies detected no association between clinicians' implicit bias and treatment recommendations [37]. A small observational study found no differences in verbal communication by race but higher nonverbal communication scores for White patients [38]. Conversely, another review noted that 5 of 6 observational and patient-reported measure-based studies found that physicians provided Black patients with less information than White patients [32]. The fact that our study found no differences in messages sent to patients by race may be because the taxonomy is based solely on the text in the message and does not leverage any nonverbal cues. Research has found evidence that nonverbal cues in text-based messages (eg, differential use of upper- and lower-case letters, spelling and grammar errors, and emoticons) can affect receivers' assessment of the senders' competence and change receivers' interpretation of the emotional intent of the message [39,40]. Thematic coding of secure messages found tone mismatches in about 16% (11/70) of the messages reviewed [41]; such mismatches could reduce patient engagement and limit patients' understanding and acceptance of any guidance provided. Comparison of message content through the more objective lens of this taxonomy coupled with a more subjective evaluation of message tone and nonverbal cues may help determine whether there are more subjective differences in message content by race or other characteristics.

Sharing medical guidance from clinicians varied by patient payer type; compared with patients with private payers, staff were more likely to send messages with medical guidance to patients with public payer types and less likely to send that content to uninsured patients. An analysis of the Medical Expenditure Panel Survey found that patients without insurance—compared with patients with public insurance—were less likely to report that their provider always listened and explained things in a way they understood [42]. Our study's findings may be an indicator from the electronic communication medium perspective of why patients without insurance might report those perceptions.

Limitations

This study was based on messages saved to patient charts because they were available for extraction at the time of this study. This means that messages sent by patients and any responses not saved to patient charts were not part of the analysis. We have no way to determine if there were trends by staff characteristics in saving messages to patient charts; therefore, we have no way to estimate whether this would further affect the associations we observed between taxa and patient and staff characteristics.

Our analysis did not include communication between patients and clinicians that occurred outside of the sampled patient-initiated threads. It is possible that a patient, for example, initiated a conversation with a clinician via phone that was concluded by a clinician-initiated thread or that a clinician responded to a patient's secure message with a phone call. We did not capture these examples and others that would fall outside of patient-initiated secure message communications. Little research has been done to explore the frequency of such cross-communications, but a small study found that approximately half of the patients' unanswered threads were resolved through other mechanisms [41]. Patient care, however, should be provided in the form needed by the patient and be responsive to patient choices and preferences [43]. If patients opt to communicate with their staff via secure messaging, it is likely that patients desire a response through that communication modality. A response through another modality may not demonstrate the best patient-centered practices. Future studies should explore whether communication occurred through other modalities and what those responses were, to better understand whether there are certain contexts when a response through an alternate modality might be appropriate, and which populations benefit from communication modality shifts.

We were also missing 5.2% (37/711) of data on staff, but that translated to 16.2% (1647/10,163) of messages not included in analyses that used staff characteristics. It is again difficult to understand the impact of this loss of data on the overall trends, but our unadjusted comparisons of the staff with unknown characteristics indicated likely within-group differences.

Finally, our cut-point values for our continuous variables were based on sample distributions. Future analyses should conduct sensitivity analyses to determine the best distribution of those cut-point values.

Conclusions

Our research presents the first analysis that associates the differences between message content and patient and staff characteristics. It demonstrates clear differences in the secure messaging content patients and staff used based not only on their respective characteristics but also those of the individuals with whom they communicated. It is important to recognize that similar to in-person communication, differences exist in communication patterns based on patient and staff characteristics. The differences we found in content use based on patient characteristics could lead to the exacerbation of health disparities when content is associated with those health outcomes. Creative technological solutions may be necessary to mitigate these differences; for example, natural language processing could be used to standardize some queries and responses and provide patients and staff with suggested text to improve communication. In the absence of a technological solution, staff and administrators should evaluate how secure messaging is used to ensure that disparities in care are not perpetuated via this communication modality.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Sampled data frequencies and odds ratios (95% CIs) estimating the association between characteristics and secure message content.
[PDF File (Adobe PDF File), 831 KB - jmir_v23i8e26650_app1.pdf]

References

1. Rathert C, Mittler JN, Banerjee S, McDaniel J. Patient-centered communication in the era of electronic health records: what does the evidence say? Patient Educ Couns 2017 Jan;100(1):50-64. [doi: 10.1016/j.pec.2016.07.031] [Medline: 27477917]
2. 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]
3. 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]
4. Shimada SL, Hogan TP, Rao SR, Allison JJ, Quill AL, Feng H, et al. Patient-provider secure messaging in VA: variations in adoption and association with urgent care utilization. Med Care 2013 Mar;51(3 Suppl 1):S21-S28. [doi: 10.1097/MLR.0b013e3182780917] [Medline: 23407007]
5. Zhou YY, Kanter MH, Wang JJ, Garrido T. Improved quality at Kaiser Permanente through e-mail between physicians and patients. Health Aff 2010 Jul;29(7):1370-1375. [doi: 10.1377/hlthaff.2010.0048] [Medline: 20606190]
6. Chung S, Panattoni L, Chi J, Palaniappan L. Can secure patient-provider messaging improve diabetes care? Diabetes Care 2017 Oct;40(10):1342-1348. [doi: 10.2337/dc17-0140] [Medline: 28807977]
7. Harris LT, Koepsell TD, Haneuse SJ, Martin DP, Ralston JD. Glycemic control associated with secure patient-provider messaging within a shared electronic medical record: a longitudinal analysis. Diabetes Care 2013 Sep;36(9):2726-2733 [FREE Full text] [doi: 10.2337/dc12-2003] [Medline: 23628618]
8. Ralston JD, Cook AJ, Anderson ML, Catz SL, Fishman PA, Carlson J, et al. Home blood pressure monitoring, secure electronic messaging and medication intensification for improving hypertension control: a mediation analysis. Appl Clin Inform 2014;5(1):232-248 [FREE Full text] [doi: 10.4338/ACI-2013-10-RA-0079] [Medline: 24734136]
9. Price-Haywood E, Luo Q, Monlezun D. Dose effect of patient-care team communication via secure portal messaging on glucose and blood pressure control. J Am Med Inform Assoc 2018 Jun 1;25(6):702-708 [FREE Full text] [doi: 10.1093/jamia/ocx161] [Medline: 29444256]
10. Harris LT, Haneuse SJ, Martin DP, Ralston JD. Diabetes quality of care and outpatient utilization associated with electronic patient-provider messaging: a cross-sectional analysis. Diabetes Care 2009 Jul;32(7):1182-1187 [FREE Full text] [doi: 10.2337/dc08-1771] [Medline: 19366959]
11. Shimada SL, Allison JJ, Rosen AK, Feng H, Houston TK. Sustained use of patient portal features and improvements in diabetes physiological measures. J Med Internet Res 2016 Jul 1;18(7):e179 [FREE Full text] [doi: 10.2196/jmir.5663] [Medline: 27369696]
12. Heisey-Grove D, McClelland LE, Rathert C, Tartaglia A, Jackson K, DeShazo J. Associations between patient health outcomes and secure message content exchanged between patients and clinicians: retrospective cohort study. J Med Internet Res 2020 Oct 29;22(10):e19477 [FREE Full text] [doi: 10.2196/19477] [Medline: 33118938]
13. Schneider MA, Howard KA. Using technology to enhance discharge teaching and improve coping for patients after stroke. J Neurosci Nurs 2017 Jun;49(3):152-156. [doi: 10.1097/JNN.0000000000000275] [Medline: 28346312]
14. Heisey-Grove DM, Carretta HJ. Disparities in secure messaging uptake between patients and physicians: longitudinal analysis of two national cross-sectional surveys. J Med Internet Res 2020 May 1;22(5):e12611 [FREE Full text] [doi: 10.2196/12611] [Medline: 32356775]
15. Sarkar U, Karter AJ, Liu JY, Adler NE, Nguyen R, López A, et al. Social disparities in internet patient portal use in diabetes: evidence that the digital divide extends beyond access. J Am Med Inform Assoc 2011 May 1;18(3):318-321 [FREE Full text] [doi: 10.1136/jamia.2010.006015] [Medline: 21262921]
16. Fareed N, Swoboda CM, Jonnalagadda P, Huerta TR. Persistent digital divide in health-related internet use among cancer survivors: findings from the health information national trends survey, 2003-2018. J Cancer Surviv 2021 Feb;15(1):87-98 [FREE Full text] [doi: 10.1007/s11764-020-00913-8] [Medline: 32671557]
17. Epstein R, Street R. Patient-centered Communication in Cancer Care: Promoting Healing and Reducing Suffering. National Cancer Institute. 2007. URL: https://cancercontrol.cancer.gov/sites/default/files/2020-06/pcc_monograph.pdf [accessed 2021-07-06]
18. Street RL, Makoul G, Arora NK, Epstein RM. How does communication heal? Pathways linking clinician-patient communication to health outcomes. Patient Educ Couns 2009 Mar;74(3):295-301. [doi: 10.1016/j.pec.2008.11.015] [Medline: 19150199]
19. Walther JB. Selective self-presentation in computer-mediated communication: hyperpersonal dimensions of technology, language, and cognition. Comput Hum Behav 2007 Sep;23(5):2538-2557. [doi: 10.1016/j.chb.2006.05.002]
20. Morrow D. Technology-based support for older adult communication in safety-critical domains. In: Ross BH, editor. Psychology of Learning and Motivation. Cambridge, Massachusetts, USA: Academic Press; 2016:285-317.

21. Gilligan T, Coyle N, Frankel RM, Berry DL, Bohlke K, Epstein RM, et al. Patient-clinician communication: American society of clinical oncology consensus guideline. J Clin Oncol 2017 Nov 1;35(31):3618-3632. [doi: 10.1200/JCO.2017.75.2311] [Medline: 28892432]
22. Byron K. Carrying too heavy a load? The communication and miscommunication of emotion by email. Acad Manag Rev 2008 Apr;33(2):309-327. [doi: 10.5465/amr.2008.31193163]
23. Trauzettel-Klosinski S, Dietz K, IReST Study Group. Standardized assessment of reading performance: the new international reading speed texts IReST. Invest Ophthalmol Vis Sci 2012 Aug 13;53(9):5452-5461. [doi: 10.1167/iovs.11-8284] [Medline: 22661485]
24. Heisey-Grove D, Rathert C, McClelland LE, Jackson K, DeShazo J. Classification of patient- and clinician-generated secure messages using a theory-based taxonomy. Health Sci Rep 2021 Jun;4(2):e295 [FREE Full text] [doi: 10.1002/hsr2.295] [Medline: 34084944]
25. 2010 Rural-Urban Commuting Area (RUCA) Codes. United States Department of Agriculture Economic Research Service. 2019 Jul 03. URL: https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/documentation/ [accessed 2019-07-14]
26. Mishel M. Uncertainty in illness. Image J Nurs Sch 1988;20(4):225-232. [doi: 10.1111/j.1547-5069.1988.tb00082.x] [Medline: 3203947]
27. Kravitz R, Bell R, Franz C. A taxonomy of requests by patients (TORP): a new system for understanding clinical negotiation in office practice. J Fam Pract 1999 Nov;48(11):872-878. [Medline: 10907624]
28. Heisey-Grove D, McClelland L, Rathert C, Tartaglia A, Jackson K, DeShazo J. Associations between patient health outcomes and secure message content exchanged between patients and clinicians: retrospective cohort study. J Med Internet Res 2020 Oct 29;22(10):e19477 [FREE Full text] [doi: 10.2196/19477] [Medline: 33118938]
29. Lafata JE, Morris HL, Dobie E, Heisler M, Werner RM, Dumenci L. Patient-reported use of collaborative goal setting and glycemic control among patients with diabetes. Patient Educ Couns 2013 Jul;92(1):94-99 [FREE Full text] [doi: 10.1016/j.pec.2013.01.016] [Medline: 23433777]
30. Sohl SJ, Borowski LA, Kent EE, Smith AW, Oakley-Girvan I, Rothman RL, et al. Cancer survivors' disclosure of complementary health approaches to physicians: the role of patient-centered communication. Cancer 2015 Mar 15;121(6):900-907 [FREE Full text] [doi: 10.1002/cncr.29138] [Medline: 25387799]
31. Fareed N, Swoboda CM, Jonnalagadda P, Walker DM, Huerta TR. Differences between races in health information seeking and trust over time: evidence from a cross-sectional, pooled analyses of hints data. Am J Health Promot 2021 Jan;35(1):84-92. [doi: 10.1177/0890117120934609] [Medline: 32588638]
32. Shen MJ, Peterson EB, Costas-Muñiz R, Hernandez MH, Jewell ST, Matsoukas K, et al. The effects of race and racial concordance on patient-physician communication: a systematic review of the literature. J Racial Ethn Health Disparities 2018 Feb;5(1):117-140 [FREE Full text] [doi: 10.1007/s40615-017-0350-4] [Medline: 28275996]
33. Heyworth L, Clark J, Marcello TB, Paquin AM, Stewart M, Archambeault C, et al. Aligning medication reconciliation and secure messaging: qualitative study of primary care providers' perspectives. J Med Internet Res 2013 Dec 2;15(12):e264 [FREE Full text] [doi: 10.2196/jmir.2793] [Medline: 24297865]
34. Ozkaynak M, Johnson S, Shimada S, Petrakis BA, Tulu B, Archambeault C, et al. Examining the multi-level fit between work and technology in a secure messaging implementation. AMIA Annu Symp Proc 2014;2014:954-962 [FREE Full text] [Medline: 25954403]
35. Wooldridge AR, Carayon P, Hoonakker P, Musa A, Bain P. Technology-Mediated Communication between Patients and Primary Care Clinicians and Staff. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 2016 Sep 15 Presented at: HFES'16; September 19-23, 2016; Washington, DC, USA p. 556-560. [doi: 10.1177/1541931213601128]
36. Pelak M, Pettit AR, Terwiesch C, Gutierrez JC, Marcus SC. Rethinking primary care visits: how much can be eliminated, delegated or performed outside of the face-to-face visit? J Eval Clin Pract 2015 Aug;21(4):591-596. [doi: 10.1111/jep.12341] [Medline: 25756943]
37. Maina IW, Belton TD, Ginzberg S, Singh A, Johnson TJ. A decade of studying implicit racial/ethnic bias in healthcare providers using the implicit association test. Soc Sci Med 2018 Feb;199:219-229. [doi: 10.1016/j.socscimed.2017.05.009] [Medline: 28532892]
38. Elliott AM, Alexander SC, Mescher CA, Mohan D, Barnato AE. Differences in physicians' verbal and nonverbal communication with black and white patients at the end of life. J Pain Symptom Manage 2016 Jan;51(1):1-8 [FREE Full text] [doi: 10.1016/j.jpainsymman.2015.07.008] [Medline: 26297851]
39. Walther JB, D'Addario K. The impacts of emoticons on message interpretation in computer-mediated communication. Soc Sci Comput Rev 2016 Aug 19;19(3):324-347. [doi: 10.1177/089443930101900307]
40. Brown S, Fuller R, Thatcher S. Impression formation and durability in mediated communication. J Assoc Inf Syst 2016 Sep;17(9):614-647. [doi: 10.17705/1jais.00436]
41. Lanham HJ, Leykum LK, Pugh JA. Examining the complexity of patient-outpatient care team secure message communication: qualitative analysis. J Med Internet Res 2018 Jul 11;20(7):e218 [FREE Full text] [doi: 10.2196/jmir.9269] [Medline: 29997107]

42. DeVoe J, Wallace L, Fryer JG. Measuring patients' perceptions of communication with healthcare providers: do differences in demographic and socioeconomic characteristics matter? Health Expect 2009 Mar;12(1):70-80 [FREE Full text] [doi: 10.1111/j.1369-7625.2008.00516.x] [Medline: 19250153]
43. 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, USA: National Academies Press (US); 2001.

Abbreviations

***ICD-10-CM*:** *International Classification of Diseases, Tenth Revision, Clinical Modification*
OR: odds ratio

Edited by R Kukafka; submitted 19.12.20; peer-reviewed by N Fareed, TV Duong; comments to author 20.02.21; revised version received 25.04.21; accepted 24.05.21; published 19.08.21.

Please cite as:
Heisey-Grove D, Rathert C, McClelland LE, Jackson K, DeShazo JP
Patient and Clinician Characteristics Associated With Secure Message Content: Retrospective Cohort Study
J Med Internet Res 2021;23(8):e26650
URL: https://www.jmir.org/2021/8/e26650
doi:10.2196/26650
PMID:34420923

©Dawn Heisey-Grove, Cheryl Rathert, Laura E McClelland, Kevin Jackson, Jonathan P DeShazo. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Artificial Intelligence–Based Prediction of Lung Cancer Risk Using Nonimaging Electronic Medical Records: Deep Learning Approach

Marvin Chia-Han Yeh[1,2], MD, PhD; Yu-Hsiang Wang[3], MD; Hsuan-Chia Yang[4,5], PhD; Kuan-Jen Bai[6,7,8], MD; Hsiao-Han Wang[1,2,4,9*], MD; Yu-Chuan (Jack) Li[1,2,4,5,9*], MD, PhD

[1]Department of Dermatology, Wan Fang Hospital, Taipei Medical University, Taipei, Taiwan
[2]Research Center of Big Data and Meta-analysis, Wan Fang Hospital, Taipei Medical University, Taipei, Taiwan
[3]School of Medicine, Taipei Medical University, Taipei, Taiwan
[4]Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan
[5]International Center for Health Information Technology, Taipei Medical University, Taipei, Taiwan
[6]Division of Pulmonary Medicine, Department of Internal Medicine, Wan Fang Hospital, Taipei Medical University, Taipei, Taiwan
[7]School of Respiratory Therapy, College of Medicine, Taipei Medical University, Taipei, Taiwan
[8]Pulmonary Research Center, Wan Fang Hospital, Taipei Medical University, Taipei, Taiwan
[9]Department of Dermatology, School of Medicine, Taipei Medical University, Taipei, Taiwan
*these authors contributed equally

Corresponding Author:
Yu-Chuan (Jack) Li, MD, PhD
Department of Dermatology
Wan Fang Hospital
Taipei Medical University
No 111, Section 3, Xinglong Road
Wenshan District
Taipei, 116
Taiwan
Phone: 886 29307930 ext 2980
Email: jaak88@gmail.com

Abstract

Background: Artificial intelligence approaches can integrate complex features and can be used to predict a patient's risk of developing lung cancer, thereby decreasing the need for unnecessary and expensive diagnostic interventions.

Objective: The aim of this study was to use electronic medical records to prescreen patients who are at risk of developing lung cancer.

Methods: We randomly selected 2 million participants from the Taiwan National Health Insurance Research Database who received care between 1999 and 2013. We built a predictive lung cancer screening model with neural networks that were trained and validated using pre-2012 data, and we tested the model prospectively on post-2012 data. An age- and gender-matched subgroup that was 10 times larger than the original lung cancer group was used to assess the predictive power of the electronic medical record. Discrimination (area under the receiver operating characteristic curve [AUC]) and calibration analyses were performed.

Results: The analysis included 11,617 patients with lung cancer and 1,423,154 control patients. The model achieved AUCs of 0.90 for the overall population and 0.87 in patients ≥55 years of age. The AUC in the matched subgroup was 0.82. The positive predictive value was highest (14.3%) among people aged ≥55 years with a pre-existing history of lung disease.

Conclusions: Our model achieved excellent performance in predicting lung cancer within 1 year and has potential to be deployed for digital patient screening. Convolution neural networks facilitate the effective use of EMRs to identify individuals at high risk for developing lung cancer.

(J Med Internet Res 2021;23(8):e26256) doi:10.2196/26256

KEYWORDS

artificial intelligence; lung cancer screening; electronic medical record

Introduction

Lung cancer is a leading cause of cancer death worldwide, and to reduce its mortality, early detection is crucial. The National Lung Cancer Screening Trial (NLST) revealed that screening with low-dose computed tomography (LDCT) can reduce the mortality associated with lung cancer by 20% [1]. Likewise, the Dutch-Belgian Randomized Lung Cancer Screening Trial (NELSON study) recently revealed that screening with LDCT resulted in a 24% decrease in the 10-year cumulative mortality for men and a 33% decrease for women [2]. Multiple organizations have recommended LDCT screening for lung cancer to be used on target populations [3,4]. Given the potential harm due to radiation exposure, false-positive results, and costs associated with LDCT, most organizations only recommend annual screening that targets high-risk individuals; this group is largely identified by epidemiological factors, including age and smoking/cessation history [5]. Furthermore, due to the potential harm associated with false-positive results, the cost-effectiveness of implementing annual LDCT screening remains controversial [6]. Multiple research groups have attempted to overcome this problem by developing risk prediction models to identify patients who might benefit from LDCT screening and to generate criteria that are superior to those introduced by the NLST and related studies [7-14]. These models frequently include self-reported information, such as family history, BMI, socioeconomic status, and smoking/cessation history, and they use conventional regression models for the final risk analysis.

In the era of digital medicine, the use of artificial intelligence has resulted in good performance for predicting image-related tasks, specifically the use of convolutional neural networks (CNNs). In lung cancer research, CNNs have been applied to LDCT and chest radiographic images to facilitate detection and classification of pulmonary nodules; these models demonstrate performance that is comparable to that achieved by human experts [15-19]. The prediction performance is largely based on high-level feature extraction and nonlinear prediction via the use of CNNs. Given proper data conversion, using CNN methodologies to generate predictions using other nonimaging medical data may be possible. Our group recently described a risk prediction model for nonmelanoma skin cancer that was generated using data extracted from electronic medical records (EMRs) [20].

In predicting lung cancer risk, the EMR should be suited to the task of identifying high-risk individuals [21]. In this study, our goal is to develop a risk model for the prediction of lung cancer using data from EMRs. As such, we applied established CNN algorithms to the large data set available in EMRs to identify important patterns associated with the development of lung cancer. In contrast with methods used for traditional regression analysis, we attempted to include evolving sequential information found in EMRs to generate our prediction model. Our goal was to generate a model that facilitated the prospective identification of individuals at higher risk for developing lung cancer; these individuals might then undergo further follow-up examinations, including LDCT. The use of a predictive model to identify individuals at high risk could serve to limit unnecessary radiation exposure and reduce costs associated with LDCT and related interventions.

Methods

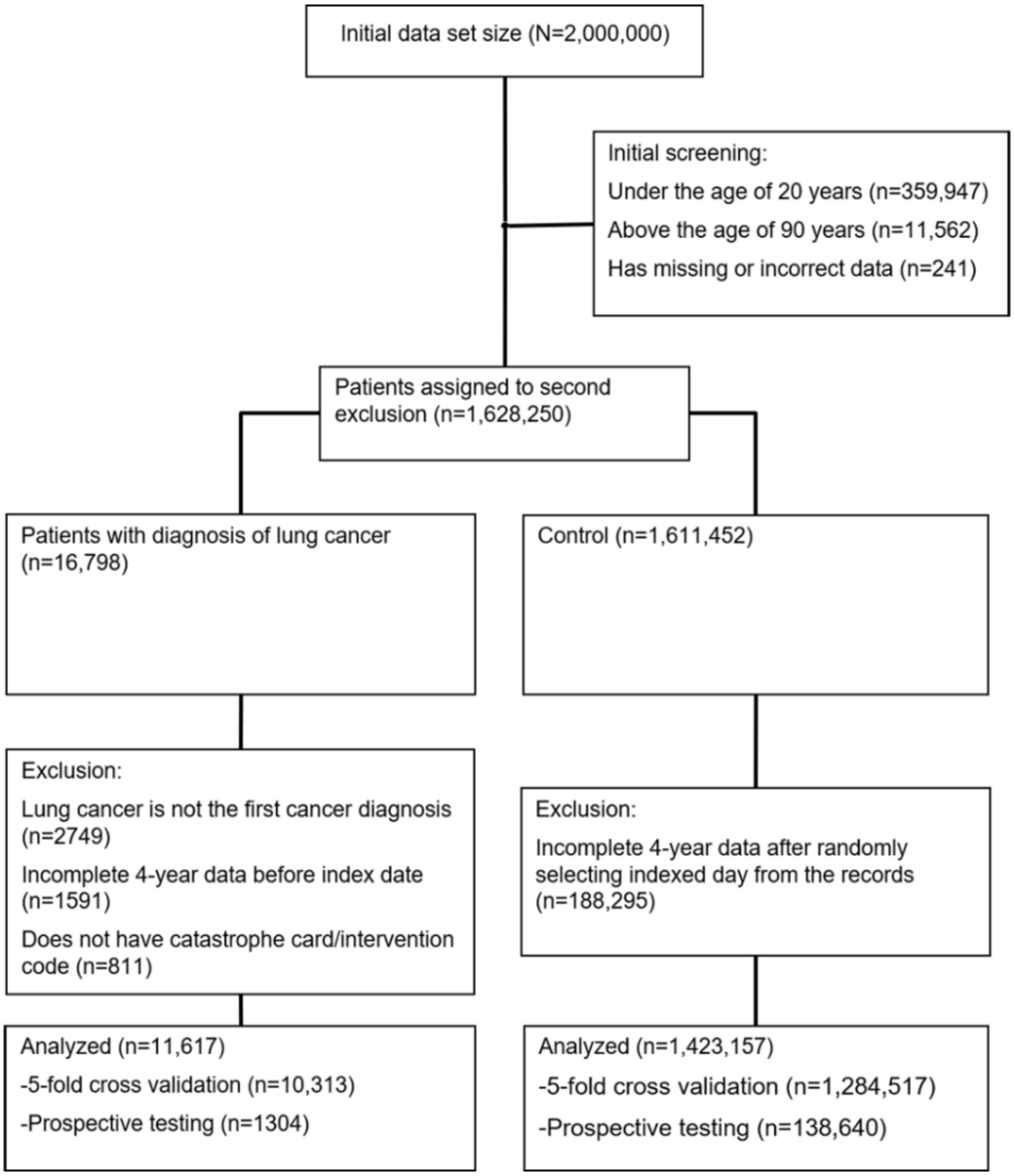
Study Population

Deidentified EMRs of 2 million patients who received care between January 01, 1999, and December 31, 2013, were initially sampled from the Taiwan National Health Insurance Research Database (NHIRD). These EMRs included the demographic information, diagnoses, and procedure codes from the *International Classification of Diseases, Ninth Revision, Clinical Modification* (*ICD-9-CM*) and prescriptions from both outpatient clinical declaration files and in-hospital declaration files. This study included participants between the ages of 20 and 90 years who had at least 4 years of medical records on file. Participants with missing data were excluded. These criteria yielded 1,628,250 EMRs with over 300 million record entries for evaluation and analysis. This study was approved by the Taipei Medical University Institutional Review Board; informed patient consent was waived, as all data were anonymous and deidentified before analysis [22].

Data Preprocessing

Previous validation studies that focused on lung cancer using the NHIRD have shown a positive predictive value (PPV) of 95% [23]. In this study, we provide further validation of the diagnosis of lung cancer using intervention codes (eg, thoracic surgery, subsequent radiotherapy, or chemotherapy) and national catastrophic illness cards (which require definite pathologic proof of a cancer diagnosis). The inclusion and exclusion criteria used in this study are indicated in Figure 1.

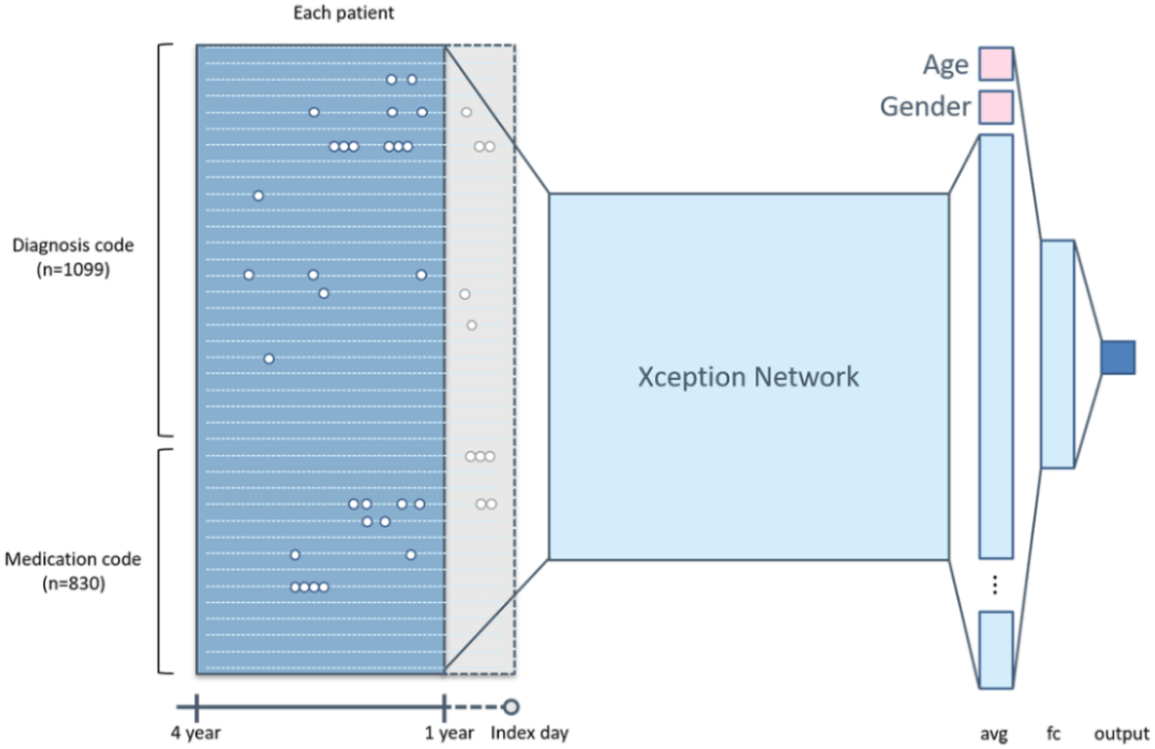
Figure 1. Inclusion and exclusion criteria for the study.


Initial data set size (N=2,000,000)
Initial screening:
Under the age of 20 years (n=359,947)
Above the age of 90 years (n=11,562)
Has missing or incorrect data (n=241)
Patients assigned to second exclusion (n=1,628,250)
Patients with diagnosis of lung cancer (n=16,798)
Control (n=1,611,452)
Exclusion:
Lung cancer is not the first cancer diagnosis (n=2749)
Incomplete 4-year data before index date (n=1591)
Does not have catastrophe card/intervention code (n=811)
Exclusion:
Incomplete 4-year data after randomly selecting indexed day from the records (n=188,295)
Analyzed (n=11,617)
-5-fold cross validation (n=10,313)
-Prospective testing (n=1304)
Analyzed (n=1,423,157)
-5-fold cross validation (n=1,284,517)
-Prospective testing (n=138,640)

The index date for patients with lung cancer was defined as the date of first diagnosis. For the control patients, the index dates were randomly selected from their medical history. *ICD-9-CM* diagnosis codes and World Health Organization-Anatomical Therapeutic Chemical (WHO-ATC) prescription codes were collected from each case for preprocessing; the date 1 year prior to the index date was used to define the prediction window. The observation window included the 3 years prior to the date included in the prediction window. Thus, we used 3 years of patient medical information to predict the risk of new-onset lung cancer at or within 1 year later (Figure 2). The *ICD-9-CM* and WHO-ATC codes were preprocessed as described in our previous study [20]. Briefly, the EMRs were classified into 1099 *ICD-9-CM* code groups and 830 WHO-ATC drug groups. Together, 1929 features were recorded weekly for 157 weeks. For each patient, the diagnoses and medications prescribed at each visit were recorded and converted to an image-like array that preserved temporal information associated with both diagnosis and medication history.

The inputs included age, gender, and an image representing the patient's 3-year history of diagnosis and medication. The image was input into Xception, a 126-layer neural network, in which feature extraction was performed. The final layer of the Xception network was connected to an average pooling layer and then connected to a fully connected layer with the patient's age and gender.

Figure 2. Visualization of the hidden layer of the model using t-stochastic neighbor embedding. Avg: average; fc: fully connected layer.



We performed 3 subgroup analyses to investigate the performance of the model in different populations. According to the age criteria used in previous trials focused on lung cancer screening [1], patients above and below 55 years of age were included among the subgroups. We also examined patients both with and without previous lung disease [24], including subgroups of patients diagnosed with asbestosis, bronchiectasis, chronic bronchitis, chronic obstructive pulmonary disease (COPD), emphysema, fibrosis, pneumonia, sarcoidosis, silicosis, and tuberculosis. Finally, to focus on the discriminative power of the diagnosis and medication without the confounding effects of age, a subgroup of age- and gender-matched controls was identified.

Model Construction and Evaluation

All patient data were split into training, validation, and testing sets based on their respective index dates. Data with index dates prior to December 31, 2012, were used for training and internal validation, and data with index dates after that date were used for prospective testing. The patients' age, gender, and image-like arrays described above were used as inputs to generate the model (Figure 2).

Lung cancer risk prediction was treated as a binary classification task using supervised learning. The model was trained to determine whether a given patient was likely to develop lung cancer within 1 year. The Xception architecture [25], which includes a 126-layer CNN-based neural network with a moderate number of parameters, was used for feature extraction. The detailed model structure is shown in Figure 2; the model construction and hyperparameters are listed in Section S1 in Multimedia Appendix 1. During training, class weights based on the population size were set to address data imbalance. To ensure the robustness of the model, a 5-fold cross validation was performed on the model. The performance of the model was assessed by its sensitivity, specificity, and area under the receiver operating characteristic curve (AUC). Model calibration was assessed using a reliability curve and the median absolute error.

To understand the model prediction, occlusion sensitivity analysis was performed by iteratively masking information from a single diagnosis or medication followed by evaluating any changes in the model prediction [26]. In addition, a dimensional reduction technique, t-distributed stochastic neighbor embedding (t-SNE), was performed on the fully connected hidden layer output of the final testing data. We randomly selected 1000 lung cancer patients and 9000 control patients for visualization. The model construction, data preprocessing, model training, and statistical processing were performed using the Python programming language, version 3.6.

Results

Baseline Demographics

A total of 11,617 lung cancer patients and 1,423,154 control patients were identified in our data set. The mean age of the lung cancer group was 66.62 years (SD 14.01); the overall data set included 856,558 (59.7%) men and 578,213 (40.3%) women. The baseline demographics of this patient cohort and the assigned subgroups are summarized in Table 1 and Tables S1-S10 in Multimedia Appendix 1.

Table 1. Demographics of the patients with lung cancer and control patients (N=1,434,771).

Group	Patients, n	Age (years), mean (SD)	Male gender, n (%)	Mean diagnosis record count (SD), n	Mean medication record count (SD), n
Whole population					
Lung cancer	11,617	66.62 (14.01)	6931 (59.7)	121.62 (113.19)	202.68 (208.97)
Control	1,423,154	44.95 (16.32)	683,375 (48.0)	66.09 (76.60)	105.99 (135.54)
Age and gender match (1:10)					
Lung cancer	11,617	66.62 (14.01)	6931 (59.7)	121.62 (113.19)	202.68 (208.97)
Control	116,169	66.62 (14.01)	69,310 (59.7)	117.99 (113.67)	190.22 (196.78)
Age ≥55 years					
Lung cancer	9261	71.99 (9.46)	5673 (61.3)	135.12 (116.31)	227.81 (218.12)
Control	385,052	66.57 (9.04)	56,730 (48.6)	114.23 (106.76)	184.50 (189.50)
Age <55 years					
Lung cancer	2356	45.50 (7.55)	1258 (53.4)	68.58 (80.42)	103.90 (126.71)
Control	1,038,102	36.93 (9.85)	496,256 (47.8)	48.23 (51.36)	76.87 (93.45)
History of lung disease[a]					
Lung cancer	3565	70.79 (12.73)	2244 (63.0)	175.12 (134.36)	297.56 (245.55)
Control	182,098	53.01 (18.09)	85,070(46.7)	125.17 (114.53)	204.85 (204.66)
No history of lung disease					
Lung cancer	8052	64.77 (14.16)	4687 (58.2)	97.94 (93.08)	160.67 (174.80)
Control	1,270,651	43.77 (15.70)	598,305 (48.2)	57.42 (64.94)	91.48 (115.23)

[a] Lung diseases included asbestosis, bronchiectasis, chronic bronchitis, chronic obstructive pulmonary disease, emphysema, fibrosis, pneumonia, sarcoidosis, silicosis, and tuberculosis. More information is provided in Table S11 in Multimedia Appendix 1.

Model Performance

For all patients, the model revealed an AUC of 0.821 when the input image-like array included sequential diagnostic information only. By contrast, the AUC was 0.894 when the input features included sequential medication information only; when the sequential diagnostic and medication information was simplified to binary variables, the model performance decreased (AUC=0.827). When both sequential diagnostic and medication information were integrated, the model reached an AUC of 0.902 on prospective testing, with a sensitivity of 0.804 and specificity of 0.837 (Table S12 in Multimedia Appendix 1). The calibration of the model showed a median expected error of 0.125; the reliability curve is shown in Figure S1 in Multimedia Appendix 1.

The model performance at different age cutoffs was then investigated. Screening using an age cutoff of 55 years revealed a superior AUC of 0.871 compared to those obtained when cutoffs of 50 or 60 years were used (0.866 and 0.863, respectively) (Table S13, Multimedia Appendix 1).

Subgroup Analysis

Analyses of the subgroups included one that was both age-and-gender-matched, those at ages above and below 55 years, and those with or without lung disease were performed. For this analysis, we identified an age- and gender-matched control subgroup that was 10 times larger than the original lung cancer subgroup. This model revealed an AUC of 0.818 (SD 0.005) with a sensitivity of 0.647 (SD 0.017) and a specificity of 0.873 (0.023 SD), as shown in Table 2 and in Table S14 in Multimedia Appendix 1. For patients above 55 years of age, the model revealed an AUC of 0.869 (SD 0.005) with a sensitivity of 0.784 (SD 0.011) and a specificity of 0.785 (SD 0.016). The PPV in this subgroup was 0.081% (SD 0.005%), and the negative predictive value was 0.993% (SD 0.000%). The performance of the model was inferior in patients below the age of 55 years; however, it still achieved an AUC of 0.815 (SD 0.007). The discriminatory powers of these models were both excellent among patients with and without a history of lung disease; the AUCs for these subgroups were 0.914 (SD 0.003) and 0.887 (SD 0.002), respectively. Among all the subgroups, the model had the weakest performance in patients below 55 years of age who had no history of lung disease; the AUC for this subgroup was only 0.797 (SD 0.008) for the one-year prospective prediction. By contrast, the model provided the strongest performance for individuals above the age of 55 years with a history of lung disease, which revealed the highest PPV of 14.3% (SD 2.3%). The model exhibited the lowest PPV of 1.0% (SD 0.2%) for individuals less than 55 years of age with no history of lung disease (Table 2). The receiver operating characteristic curves associated with each of these subgroups are summarized in sections S2.1-S2.9 in Multimedia Appendix 1.

Table 2. Discrimination performance (testing set) of the model in the subgroups.

Subgroup	Lung cancer group, n	Control, n	Testing AUC[a] (SD)	Testing sensitivity (SD)	Testing specificity (SD)	PPV[b] (SD), %	NPV[c] (SD), %
Whole population	1304	138,640	0.898 (0.002)	0.805 (0.015)	0.825 (0.018)	4.2 (0.3)	99.8 (0)
Matching age and gender	1304	13,040	0.818 (0.005)	0.647 (0.017)	0.873 (0.023)	34.6 (0.4)	96.0 (0.1)
Age ≥55 years	1046	43,328	0.869 (0.002)	0.784 (0.011)	0.785 (0.016)	8.1 (0.5)	99.3 (0)
Age <55 years	258	95,312	0.815 (0.007)	0.620 (0.080)	0.838 (0.054)	1.1 (0.2)	99.9 (0)
History of lung disease	361	16,596	*0.914 (0.003)*[d]	0.829 (0.021)	0.816 (0.021)	9.0 (0.8)	0.995 (0.1)
No history of lung disease	943	122,044	0.887 (0.002)	0.781 (0.025)	0.827 (0.026)	3.4 (0.5)	99.8 (0.0)
Age ≥55 years with history of lung disease	318	8184	0.875 (0.005)	0.755 (0.047)	0.819 (0.044)	*14.3 (2.3)*	98.9 (0.2)
Age ≥55 years with no history of lung disease	728	35,144	0.865 (0.003)	0.775 (0.019)	0.786 (0.018)	7.0 (0.4)	99.4 (0.0)
Age <55 years with history of lung disease	43	8,412	0.909 (0.006)	0.777 (0.054)	0.891 (0.036)	3.8 (1.0)	99.9 (0.0)
Age <55 years with no history of lung disease	215	86,900	0.797 (0.008)	0.533 (0.048)	0.865 (0.026)	*1.0 (0.2)*	99.9 (0.0)

[a] AUC: area under the curve.

[b] PPV: positive predictive value.

[c] NPV: negative predictive value.

[d] Italic text indicates the best performance for the parameter.

Table 3 summarizes the age, gender, diagnosis, and medications associated with both the correctly and incorrectly classified groups from the testing data set. The mean age of the true-positive group was similar to that of the false-positive group and somewhat greater than that of the false-negative group. This tendency was also observed in other subgroups; overall, our results suggest that age and sex are important predictive factors. This is consistent with the t-SNE analysis, in which patients with lung cancer and control patients over 55 years of age were clustered centrally, as compared to the other patients, who were located at the periphery (Figure 3).

The model's hidden layer outputs of 1000 patients with cancer (red dots) and 9000 control patients (green dots) were visualized using t-SNE (Figure 3). Dark green and red represent old age control patients and patients with cancer, respectively. As shown in the left image, most patients with cancer can be clustered away from the control patients. Some dark red dots are mixed with dark green dots in the upper area. These are the patients that were wrongly predicted to be controls by the model. The center images shows that patients aged ≥55 years were clustered in the center of the graph, with the patients with cancer were successfully clustered in the tip area. The right image shows that patients aged <55 years were clustered at the periphery of the graph. Some patients with cancer were also clustered in the tip area, whereas the others were scattered with the control patients.

Occlusion sensitivity analysis further revealed that the specific diagnosis and medication factors were associated with an increased risk of developing lung cancer. Interestingly, "other noninfectious gastroenteritis and colitis" and "other agents for local oral treatment" were associated with the highest risks of developing lung cancer with respect to patient diagnosis and medication, respectively. The top 20 factors identified in the analysis are summarized in Table 4.

Table 3. Prediction analysis of the prospective testing data set (N=139,944).

Group	Patients, n	Age (years), mean (SD)	Male gender, n (%)	Mean diagnosis count (SD), n	Mean medication count (SD), n
All patients					
True positive	1052	69.91 (11.58)	617 (58.65)	141.75 (113.31)	210.7 (186.32)
False positive	22,624	69.19 (12.48)	12,641 (55.87)	114.96 (111.04)	159.14 (171.74)
True negative	116,016	41.94 (13.14)	53,671 (46.26)	63.08 (67.53)	81.46 (101.84)
False negative	252	50.96 (10.79)	134 (53.17)	81.37 (95.67)	104.03 (139.98)
Patients aged ≥55 years					
True positive	851	72.86 (9.25)	510 (59.93)	146.32 (110.84)	217.88 (181.04)
False positive	10,989	74.88 (9.66)	6640 (60.42)	124.11 (119.27)	170.8 (179.15)
True negative	32,339	63.28 (6.58)	13,871 (42.89)	110.24 (97.26)	152.69 (154.96)
False negative	195	64.62 (6.63)	106 (54.36)	125.98 (132.09)	185.08 (216.55)
Patients aged <55 years					
True positive	209	47.87 (6.07)	113 (54.07)	83.3 (87.98)	106.48 (128.64)
False positive	32,765	46.78 (6.58)	18,422 (56.22)	59.4 (63.22)	74.38 (92.27)
True negative	62,547	32.45 (7.43)	27,379 (43.77)	48.67 (48.88)	60.74 (71.36)
False negative	49	36.22 (5.82)	22 (44.90)	63.98 (63.75)	83.88 (115.66)
Patients with a history of lung disease					
True positive	300	72.86 (11.18)	182 (60.67)	184.91 (118.07)	278.71 (194.81)
False positive	2791	75.41 (11.97)	1750 (62.70)	180.66 (140.56)	253.68 (214.05)
True negative	13,805	49.34 (15.6)	5876 (42.56)	119.33 (102.8)	162.24 (162.85)
False negative	61	61.41 (12.11)	34(55.74)	171.72 (155.81)	246.79 (226.86)
Patients with no history of lung disease					
True positive	757	68.45 (11.4)	442 (58.39)	120.97 (104.28)	177.03 (172.5)
False positive	23,328	66.54 (12.25)	12,881 (55.22)	95.23 (94.24)	130.24 (146.34)
True negative	98,716	40.39 (12.27)	45,805 (46.40)	56.19 (59.51)	71.56 (88.63)
False negative	186	48.19 (10.32)	93 (50.00)	65.08 (66.98)	81.69 (101.83)
Patients aged ≥55 years with a history of lung disease					
True positive	255	74.89 (9.03)	160 (62.75)	188.33 (119.58)	284.4 (193.99)
False positive	1778	78.53 (9.16)	1205 (67.77)	188.16 (142.99)	263 (215.97)
True negative	6406	66.38 (7.88)	2669 (41.66)	169.82 (121.41)	239.26 (195.71)
False negative	63	70.44 (7.81)	35 (55.56)	203.87 (148.87)	308.17 (221.29)
Patients aged ≥55 years with no history of lung disease					
True positive	587	71.76 (9.24)	347(59.11)	126.04 (102.89)	185.01 (166.72)
False positive	8958	73.86 (9.69)	5,281(58.95)	104.85 (103.3)	142.56 (154.72)
True negative	26,186	62.73 (6.27)	11,356(43.37)	98.04 (87.47)	135.09 (139.76)
False negative	141	63.47 (6.25)	74(52.48)	100.89 (103.77)	148.73 (195.18)
Patients aged <55 years with lung diseases					
True positive	37	48.89 (6.08)	18 (48.65)	120.46 (100.27)	157.62 (173.25)
False positive	1080	46.56 (7.56)	653 (60.46)	85.56 (72.24)	109.78 (108.74)
True negative	7332	37.7 (9.58)	3099 (42.27)	86.84 (75.16)	113.06 (116.51)
False negative	6	43.33 (9.24)	3 (50.00)	103.67 (98.36)	149.83 (152.85)
Patients aged <55 years with no history of lung disease					

Group	Patients, n	Age (years), mean (SD)	Male gender, n (%)	Mean diagnosis count (SD), n	Mean medication count (SD), n
True positive	172	47.55 (6.07)	95(55.23)	74.94 (83.33)	94.44 (114.72)
False positive	30,982	46.56 (6.56)	17,478(56.41)	55.1 (58.63)	68.47 (84.96)
True negative	55,918	32.06 (7.25)	24,571(43.94)	45.68 (45.68)	56.64 (65.81)
False negative	43	35.65 (5.54)	19(44.19)	59.88 (56.98)	78.84 (108.63)

Figure 3. Visualization of the hidden layer of the model using t-stochastic neighbor embedding.

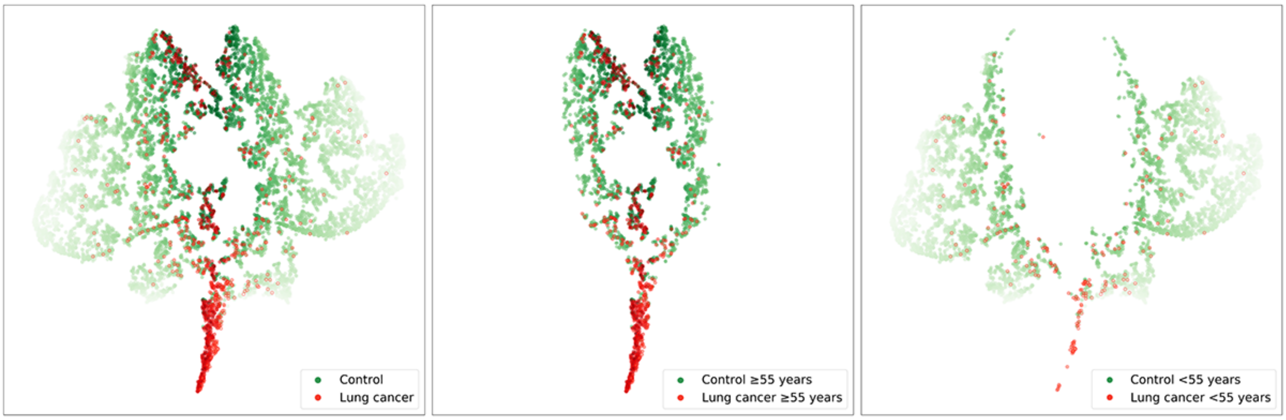


Table 4. Top 20 factors related to lung cancer learned by the model.

Rank	Factor	Lung cancer risk increase (%), mean (SD)
1	Other noninfectious gastroenteritis and colitis	1.85 (1.01)
2	Other congenital anomalies of the circulatory system	1.84 (2.21)
3	Other agents for local oral treatment	1.76 (1.02)
4	Antidotes	1.69 (1.55)
5	Postinflammatory pulmonary fibrosis	1.69 (1.43)
6	Metronidazole	1.69 (1.29)
7	Acariasis	1.65 (1.73)
8	Antiviral drugs	1.57 (1.03)
9	Orchitis and epididymitis	1.57 (1.48)
10	Pneumococcal pneumonia	1.52 (0.93)
11	Buflomedil	1.44 (1.76)
12	Danazol	1.42 (1.41)
13	Calcineurin inhibitors	1.42 (1.29)
14	Other disorders of the urethra and urinary tract	1.37 (1.34)
15	Angina pectoris	1.35 (1.44)
16	Other nonorganic psychoses	1.35 (1.99)
17	Respiratory conditions due to other and unspecified external agents	1.33 (1.33)
18	Open wound of back	1.33 (2.46)
19	Hydrazinophthalazine derivatives	1.31 (1.57)
20	Insulin	1.30 (1.51)

Discussion

Principal Findings

In this study, we explored the possibility of predicting lung cancer using a CNN with diagnosis and medication history extracted from EMRs as a data source. Unlike other proposed lung cancer risk models, our model does not rely on self-reported parameters such as smoking/cessation history, family history, socioeconomic status, or BMI. This model could be readily deployed as a means to evaluate centralized health care databases and perform efficient population-based screening. Such an approach has potential to improve the accuracy of current screening methods, as it can identify those most likely to benefit from interventions [21]. In addition, we attempted to include time-related sequential information as reflected in the medical histories as a means to evaluate lung cancer risk. This approach is different from those used in traditional regression analysis, in which personal history is often simplified and limited to binary or categorical variables. We found that the integration of temporal aspects resulted in improvements in the model performance (Table S12 in Multimedia Appendix 1). The capacity for complex integration of multiple variables is one of the strengths of deep neural networks. To generate this model, we used an established computer vision model (Xception) to extract high-level features from the array representing individual clinical case histories; this ensured that the high-level features associated with the clinical information were effectively extracted for risk prediction.

Related Work

Lung cancer prediction models are under investigation with the goal of identifying high-risk populations that might benefit from LDCT screening. A variety of parameters have been used for prediction, including epidemiologic factors (eg, socioeconomic status, BMI, and smoking history), clinical history (eg, family history and individual history of lung disease history), and results of clinical examinations (eg, blood tests, genetic analysis, and imaging results). The PLCOm2012 model is the most widely validated, with AUCs of 0.78 to 0.82 [27-30]. Likewise, the Bach model exhibited AUCs of 0.66 to 0.75 on external validation [5,31]. Other models include the Haggart model, which exhibited AUCs of 0.71 to 0.84 [5,9], the Liverpool Lung Project model, with AUCs of 0.67 to 0.82 [32], and the Lung Cancer Risk Assessment Tool, which achieved AUCs of 0.77 to 0.78 [5,33]. Some models used information extracted from patient EMRs. The model proposed by Iyen-Omofoman et al [10] used lung-associated clinical symptoms and social-epidemiologic factors from a general practice database, and they achieved an AUC of 0.88; likewise, Wang et al [13] included 33,788 clinical features from clinical histories and laboratory tests evaluated in an extreme gradient boosting (XGBoost) model to achieve an AUC of 0.88. With these previous studies in mind, our model featured a deep learning approach and achieved a prospective prediction AUC of 0.87 in patients older than 55 years and 0.90 for the entire patient cohort. It is possible to test other machine learning models (eg, support vector machine or random forest) on our data set. However, this study serves as a proof of concept of using CNN with nonimaging medical records. Comparing the performance of this model to that of different machine learning models of practical interest would be an interesting approach for future studies.

We recognize that direct comparisons between models may not be fully appropriate, as the target populations and predicted outcomes can vary. Previous reports suggested that the performance of models is inflated when nonsmokers and younger subjects (<55 years of age) are included in the study groups [34]. Our findings confirm this point, as can be observed from the higher AUCs associated with the younger age cutoffs (Table S3, Multimedia Appendix 1). Although our data set did not directly include reports of smoking history or cessation, we did include a history of lung diseases (eg, chronic bronchitis, COPD, and emphysema) among our parameters; these could easily be considered as surrogate factors for smoking history. Further analysis of this patient subgroup may help us understand and mitigate the possibility of performance inflation.

In the original NLST trial, the PPV for the LDCT was determined to be 3.4% [1]. The high false-positive rate associated with this intervention remains a major concern with respect to LDCT screening. In this study, the highest PPV (14.5%) was observed in patients ≥55 years of age with a history of lung disease. As noted above, an increase in cancer diagnoses might be expected in this patient subgroup, as a history of lung disease may be a direct consequence of smoking. As such, this finding suggested that individuals in this subgroup are suitable candidates for model prescreening in an effort to avoid unnecessary radiation exposure and costs associated with LDCT. In addition, we found that the 55-year age cutoff selected in the original NLST trial was also appropriate for our model, as the predictive performance was higher with this age cutoff compared to that observed at cutoffs at age 50 or 60 years (Table S3, Multimedia Appendix 1).

Predictive Factor Analysis

The inclusion of an age- and gender-matched subgroup was necessary to explore the roles of clinical diagnosis and medication history in the predictions generated by our model; evaluation of this subgroup prevented the confounding effects of age and its correlations to clinical history (eg, older people are typically prescribed more chronic disease-related medications). With this consideration, our model achieved an AUC of 0.818. These findings can be compared to the model proposed by Spitz et al [12], which included gender-, age-, and smoking status–matched patients and achieved an AUC of 0.63 in former smokers. Although the models generated from matched populations tended to display weaker performance than those from nonmatched populations and may not be clinically useful, this result provided us with a more clear-cut evaluation of the specific parameters included in this model. Taken together, our findings suggest that our model is capable of identifying factors that are useful for predicting lung cancer using clinical information available 1 year before the clinical diagnosis is made.

Our model demonstrated the worst performance in young patients without pre-existing lung diseases. This finding suggests that identifying high-risk patients among young and asymptomatic patients is still the most challenging task. Further studies are required to assess the performance of the model in patients with different staging. One of the major concerns with respect to the use of lung cancer prediction models is that they tend to select individuals who are older and who have multiple comorbidities [35], thus reducing the overall benefit gained from the screening process [36]. This tendency was also observed in our model. This phenomenon cannot be fully avoided, as it simply reflects the high percentage of older patients in the population who are diagnosed with lung cancer. However, when focused on patients younger than 55 years of age, our model maintained excellent discriminative power (the AUC was 0.82, with a mean age of true positives of 47.8 years). With the current model, the inclusion of younger individuals remains possible; multiple age-stratified thresholds for lung cancer risk could further optimize the clinical benefits of the predictions from this model.

Although deep learning is often considered a “black box,” and it is often challenging to explain the reasoning behind the outcomes, our study used t-SNE and occlusion sensitivity analysis to identify the most critical of the contributing parameters. Our occlusion sensitivity analysis revealed that many of the important factors were those associated with a history of preexisting lung conditions (eg, postinflammatory pulmonary fibrosis and pneumococcal pneumonia) and medications used to treat smoking-related diseases (eg, buflomedil for peripheral arterial disease and angina pectoris, and insulin for insulin resistance of diabetes mellitus) with increased cancer risk (eg, congenital anomalies of the circulatory system [37] and periodontal conditions [38]), and paraneoplastic phenomena (eg, noninfectious gastroenteritis and colitis [39]). This information must be interpreted carefully, as these findings do not imply a causal relationship. For example, the model may predict an increased likelihood of future lung cancer in patients with pre-existing lung disease simply because these patients receive frequent medical attention; thus, there is a higher likelihood that cancer will be detected incidentally. In addition, the sensitivity analysis in this study is only capable of evaluating one factor at a time; this is a major limitation of the explainability of the model, given the fact that our model was designed to integrate complex, high-level features. Finally, we could not explain some of the predictive features identified by this model, such as the associations with terms including *antidote*, *orchitis*, and *epididymitis*. More studies will be required to decode the findings from the CNN and to elucidate the interactions between age, sex, previous diagnoses, and medications.

Although our model achieved excellent discriminative performance, poor calibration was noted, together with the fact that direct numeric output would overestimate the actual risk. This is a known phenomenon associated with modern neural networks [40]. Unlike the traditional logistic regression models, which perform well in calibration because they directly minimize the loss of calibration, modern neural networks tend to perform suboptimally in this regard. This is likely due to the regularization methods (eg, dropout and batch normalization) and the multiple deep layers applied as components of the model architecture [40]. In our study, poor calibration did not limit the use of the model, as individuals were selected based on a predefined threshold identified in the validation data set rather than on the numerical output of the model. As a result, the increased rates reported in Table 4 do not represent the actual cancer risk.

Our model used nonimaging medical information from EMRs; however, we still used CNN as the model backbone. The study design and aims are different from other lung cancer studies that used CNN to analyze computed tomography (CT) scans and determine if a pulmonary nodule is malignant. Their models were used to automatically identify suspicious nodules from CT scans, which were already present, whereas our model attempted to identify patients with high risk of developing lung cancer in the future.

Limitations

There are several limitations to this study. First, the data collection was limited to the NHIRD database of Taiwan; the patient records do not include tissue histology or lung cancer staging data. Patients with small cell lung cancer and mutation-rich non–small cell lung cancer (eg, epidermal growth factor receptor, anaplastic lymphoma kinase, ROS-1) could not be separated. These specific types may have different disease courses and risk factors; therefore, they were usually not included in the traditional screening, and the benefit of receiving screening is undetermined. Our subgroup analysis did include only patients with pre-existing lung diseases, but this did not mitigate the issue entirely. Similarly, the NHIRD database does not include information on patients’ lifestyles or any genetic or laboratory data. A subgroup analysis of patients with lung cancer based on tissue histology and staging might help to develop a prediction model that was tailored to different risk groups. Second, the data set did not contain any information on smoking status, which is clearly an important risk factor associated with lung cancer development. This limitation restricted the external validation and the comparisons that could be made between our model and those described in earlier published studies. The authors believe that self-reported information, such as family history, smoking/cessation history, and duration of symptoms, are valuable pieces of information for lung cancer prediction that are very important and can further improve prediction accuracy. In our study, a history of lung diseases (eg, COPD and emphysema) was used as a proxy for a smoking history; our model performed with excellent discriminative power with respect to this subgroup. Finally, the NHIRD includes primarily Taiwanese people; as such, the target population was fairly homogeneous, with limited ethnic diversity. The identified risk factors may not apply to other populations with other ethnicities. Nonetheless, the methodology used here could be easily applied to other medical databases with more diverse patient populations.

Conclusion

Our CNN model exhibited robust performance with respect to the 1-year prospective prediction of the risk of developing lung cancer. As our model included sequential data on clinical diagnoses and medication history, it was capable of capturing features associated with evolving clinical conditions and as such was able to identify patients at higher risk of developing lung cancer. With appropriate ethical regulation, this model may be deployed as a means to analyze medical databases, thus paving the way for efficient population-based screening and digital precision medicine. A future randomized controlled trial will be required to explore the clinical benefit of this model in diverse populations.

Acknowledgments

This research was funded in part by Ministry of Education (MOE) grants MOE 109-6604-001-400 and DP2-110-21121-01-A-01.

Authors' Contributions

MCHY contributed to the data analysis, model construction, interpretation of results, drafting of the manuscript, and literature review. YHW and HCY contributed to the data curation and data preprocessing. KJB contributed to the investigation and the interpretation of the results. HHW contributed to the interpretation of results, conceptualization, supervision, and manuscript editing. YCL contributed to the conceptualization, supervision, manuscript editing, and interpretation of the results. HHW and YCL contributed equally to this article. The corresponding author, YCL, 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 planned (and, if relevant, registered) have been explained.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables and figures.
[DOCX File , 770 KB - jmir_v23i8e26256_app1.docx]

References

1. National Lung Screening Trial Research Team, Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med 2011 Aug 04;365(5):395-409 [FREE Full text] [doi: 10.1056/NEJMoa1102873] [Medline: 21714641]
2. de Koning HJ, van der Aalst CM, de Jong PA, Scholten ET, Nackaerts K, Heuvelmans MA, et al. Reduced lung-cancer mortality with volume CT screening in a randomized trial. N Engl J Med 2020 Feb 06;382(6):503-513. [doi: 10.1056/NEJMoa1911793] [Medline: 31995683]
3. Moyer VA, U.S. Preventive Services Task Force. Screening for lung cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med 2014 Mar 04;160(5):330-338 [FREE Full text] [doi: 10.7326/M13-2771] [Medline: 24378917]
4. Wood DE, Kazerooni EA, Baum SL, Eapen GA, Ettinger DS, Hou L, et al. Lung cancer screening, version 3.2018, NCCN clinical practice guidelines in oncology. J Natl Compr Canc Netw 2018 Apr;16(4):412-441 [FREE Full text] [doi: 10.6004/jnccn.2018.0020] [Medline: 29632061]
5. Katki HA, Kovalchik SA, Petito LC, Cheung LC, Jacobs E, Jemal A, et al. Implications of nine risk prediction models for selecting ever-smokers for computed tomography lung cancer screening. Ann Intern Med 2018 Jul 03;169(1):10-19 [FREE Full text] [doi: 10.7326/M17-2701] [Medline: 29800127]
6. Ruano-Ravina A, Pérez-Ríos M, Casàn-Clará P, Provencio-Pulla M. Low-dose CT for lung cancer screening. Lancet Oncol 2018 Mar;19(3):e131-e132. [doi: 10.1016/S1470-2045(18)30121-9] [Medline: 29508756]
7. Bach PB, Kattan MW, Thornquist MD, Kris MG, Tate RC, Barnett MJ, et al. Variations in lung cancer risk among smokers. J Natl Cancer Inst 2003 Mar 19;95(6):470-478. [doi: 10.1093/jnci/95.6.470] [Medline: 12644540]
8. Cassidy A, Myles JP, van Tongeren M, Page RD, Liloglou T, Duffy SW, et al. The LLP risk model: an individual risk prediction model for lung cancer. Br J Cancer 2008 Jan 29;98(2):270-276 [FREE Full text] [doi: 10.1038/sj.bjc.6604158] [Medline: 18087271]
9. Hoggart C, Brennan P, Tjonneland A, Vogel U, Overvad K, Østergaard JN, et al. A risk model for lung cancer incidence. Cancer Prev Res (Phila) 2012 Jun;5(6):834-846 [FREE Full text] [doi: 10.1158/1940-6207.CAPR-11-0237] [Medline: 22496387]
10. Iyen-Omofoman B, Tata LJ, Baldwin DR, Smith CJP, Hubbard RB. Using socio-demographic and early clinical features in general practice to identify people with lung cancer earlier. Thorax 2013 May;68(5):451-459. [doi: 10.1136/thoraxjnl-2012-202348] [Medline: 23321602]
11. Marcus MW, Chen Y, Raji OY, Duffy SW, Field JK. LLPi: Liverpool Lung Project risk prediction model for lung cancer incidence. Cancer Prev Res (Phila) 2015 Jun;8(6):570-575 [FREE Full text] [doi: 10.1158/1940-6207.CAPR-14-0438] [Medline: 25873368]

12. Spitz MR, Hong WK, Amos CI, Wu X, Schabath MB, Dong Q, et al. A risk model for prediction of lung cancer. J Natl Cancer Inst 2007 May 02;99(9):715-726. [doi: 10.1093/jnci/djk153] [Medline: 17470739]
13. Wang X, Zhang Y, Hao S, Zheng L, Liao J, Ye C, et al. Prediction of the 1-year risk of incident lung cancer: prospective study using electronic health records from the state of Maine. J Med Internet Res 2019 May 16;21(5):e13260 [FREE Full text] [doi: 10.2196/13260] [Medline: 31099339]
14. Wilson DO, Weissfeld J. A simple model for predicting lung cancer occurrence in a lung cancer screening program: The Pittsburgh Predictor. Lung Cancer 2015 Jul;89(1):31-37 [FREE Full text] [doi: 10.1016/j.lungcan.2015.03.021] [Medline: 25863905]
15. Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nat Med 2019 Jun;25(6):954-961. [doi: 10.1038/s41591-019-0447-x] [Medline: 31110349]
16. Jena S, George S. Morphological feature extraction and KNG - CNN classification of CT images for early lung cancer detection. Int J Imaging Syst Technol 2020 Jun 09;30(4):1324-1336 [FREE Full text] [doi: 10.1002/ima.22445]
17. Priya MMA, Jawhar SJ. Advanced lung cancer classification approach adopting modified graph clustering and whale optimisation - based feature selection technique accompanied by a hybrid ensemble classifier. IET Image Process 2020 Jul 08;14(10):2204-2215 [FREE Full text] [doi: 10.1049/iet-ipr.2019.0178]
18. Manikandan T, Bharathi N. Lung cancer detection using fuzzy auto-seed cluster means morphological segmentation and SVM classifier. J Med Syst 2016 Jul;40(7):181. [doi: 10.1007/s10916-016-0539-9] [Medline: 27299354]
19. Gao N, Tian S, Li X, Huang J, Wang J, Chen S, et al. Three-dimensional texture feature analysis of pulmonary nodules in CT images: lung cancer predictive models based on Support Vector Machine Classifier. J Digit Imaging 2020 Apr;33(2):414-422 [FREE Full text] [doi: 10.1007/s10278-019-00238-8] [Medline: 31529236]
20. Wang H, Wang Y, Liang C, Li Y. Assessment of deep learning using nonimaging information and sequential medical records to develop a prediction model for nonmelanoma skin cancer. JAMA Dermatol 2019 Nov 01;155(11):1277-1283 [FREE Full text] [doi: 10.1001/jamadermatol.2019.2335] [Medline: 31483437]
21. Wender RC, Brawley OW, Fedewa SA, Gansler T, Smith RA. A blueprint for cancer screening and early detection: advancing screening's contribution to cancer control. CA Cancer J Clin 2019 Jan;69(1):50-79 [FREE Full text] [doi: 10.3322/caac.21550] [Medline: 30452086]
22. 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]
23. Kao W, Hong J, See L, Yu H, Hsu J, Chou I, et al. Validity of cancer diagnosis in the National Health Insurance database compared with the linked National Cancer Registry in Taiwan. Pharmacoepidemiol Drug Saf 2018 Oct;27(10):1060-1066. [doi: 10.1002/pds.4267] [Medline: 28815803]
24. Kovalchik SA, Tammemagi M, Berg CD, Caporaso NE, Riley TL, Korch M, et al. Targeting of low-dose CT screening according to the risk of lung-cancer death. N Engl J Med 2013 Jul 18;369(3):245-254 [FREE Full text] [doi: 10.1056/NEJMoa1301851] [Medline: 23863051]
25. Chollet F. Xception: deep learning with depthwise separable convolutions. 2017 Apr 04 Presented at: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); July 22-25, 2017; Honolulu, HI. [doi: 10.1109/cvpr.2017.195]
26. Zeiler M, Fergus R. Visualizing and understanding convolutional networks. In: Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8689. 2014 Presented at: European Conference on Computer Vision 2014; September 6-12, 2014; Zurich, Switzerland p. 818-833. [doi: 10.1007/978-3-319-10590-1_53]
27. Li K, Hüsing A, Sookthai D, Bergmann M, Boeing H, Becker N, et al. Selecting high-risk individuals for lung cancer screening: a prospective evaluation of existing risk models and eligibility criteria in the German EPIC cohort. Cancer Prev Res (Phila) 2015 Sep;8(9):777-785 [FREE Full text] [doi: 10.1158/1940-6207.CAPR-14-0424] [Medline: 26076698]
28. Tammemagi MC, Schmidt H, Martel S, McWilliams A, Goffin JR, Johnston MR, PanCan Study Team. Participant selection for lung cancer screening by risk modelling (the Pan-Canadian Early Detection of Lung Cancer [PanCan] study): a single-arm, prospective study. Lancet Oncol 2017 Nov;18(11):1523-1531. [doi: 10.1016/S1470-2045(17)30597-1] [Medline: 29055736]
29. Weber M, Yap S, Goldsbury D, Manners D, Tammemagi M, Marshall H, et al. Identifying high risk individuals for targeted lung cancer screening: independent validation of the PLCO risk prediction tool. Int J Cancer 2017 Jul 15;141(2):242-253 [FREE Full text] [doi: 10.1002/ijc.30673] [Medline: 28249359]
30. Tammemägi MC, Katki HA, Hocking WG, Church TR, Caporaso N, Kvale PA, et al. Selection criteria for lung-cancer screening. N Engl J Med 2013 Feb 21;368(8):728-736 [FREE Full text] [doi: 10.1056/NEJMoa1211776] [Medline: 23425165]
31. D'Amelio AM, Cassidy A, Asomaning K, Raji OY, Duffy SW, Field JK, et al. Comparison of discriminatory power and accuracy of three lung cancer risk models. Br J Cancer 2010 Jul 27;103(3):423-429 [FREE Full text] [doi: 10.1038/sj.bjc.6605759] [Medline: 20588271]
32. Raji OY, Duffy SW, Agbaje OF, Baker SG, Christiani DC, Cassidy A, et al. Predictive accuracy of the Liverpool Lung Project risk model for stratifying patients for computed tomography screening for lung cancer: a case-control and cohort validation study. Ann Intern Med 2012 Aug 21;157(4):242-250 [FREE Full text] [doi: 10.7326/0003-4819-157-4-201208210-00004] [Medline: 22910935]

33. Katki HA, Kovalchik SA, Berg CD, Cheung LC, Chaturvedi AK. Development and validation of risk models to select ever-smokers for CT lung cancer screening. JAMA 2016 Jun 07;315(21):2300-2311 [FREE Full text] [doi: 10.1001/jama.2016.6255] [Medline: 27179989]
34. Tammemägi MC. Application of risk prediction models to lung cancer screening: a review. J Thorac Imaging 2015 Mar;30(2):88-100. [doi: 10.1097/RTI.0000000000000142] [Medline: 25692785]
35. Mazzone PJ, Silvestri GA, Patel S, Kanne JP, Kinsinger LS, Wiener RS, et al. Screening for lung cancer: CHEST guideline and expert panel report. Chest 2018 Apr;153(4):954-985. [doi: 10.1016/j.chest.2018.01.016] [Medline: 29374513]
36. Cheung LC, Berg CD, Castle PE, Katki HA, Chaturvedi AK. Life-gained-based versus risk-based selection of smokers for lung cancer screening. Ann Intern Med 2019 Nov 05;171(9):623-632 [FREE Full text] [doi: 10.7326/M19-1263] [Medline: 31634914]
37. Mandalenakis Z, Karazisi C, Skoglund K, Rosengren A, Lappas G, Eriksson P, et al. Risk of cancer among children and young adults with congenital heart disease compared with healthy controls. JAMA Netw Open 2019 Jul 03;2(7):e196762 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.6762] [Medline: 31276179]
38. Mai X, LaMonte MJ, Hovey KM, Nwizu N, Freudenheim JL, Tezal M, et al. History of periodontal disease diagnosis and lung cancer incidence in the Women's Health Initiative Observational Study. Cancer Causes Control 2014 Aug;25(8):1045-1053 [FREE Full text] [doi: 10.1007/s10552-014-0405-3] [Medline: 24913780]
39. Schmidt-Hieber M, Bierwirth J, Buchheidt D, Cornely OA, Hentrich M, Maschmeyer G, AGIHO Working Group. Diagnosis and management of gastrointestinal complications in adult cancer patients: 2017 updated evidence-based guidelines of the Infectious Diseases Working Party (AGIHO) of the German Society of Hematology and Medical Oncology (DGHO). Ann Hematol 2018 Jan;97(1):31-49 [FREE Full text] [doi: 10.1007/s00277-017-3183-7] [Medline: 29177551]
40. Guo C, Pleiss G, Sun Y, Weinberger KQ. On calibration of modern neural networks. In: Proceedings of the 34th International Conference on Machine Learning. 2017 Presented at: 34th International Conference on Machine Learning; August 6-11, 2017; Sydney, Australia p. 1321-1330.

Abbreviations

ATC: Anatomical Therapeutic Chemical
AUC: area under the receiver operating characteristic curve
CNN: convolutional neural network
COPD: chronic obstructive pulmonary disease
CT: computed tomography
EMR: electronic medical record
ICD-9-CM: International Classification of Diseases, Ninth Revision, Clinical Modification
LDCT: low-dose computed tomography
MOE: Ministry of Education
NHIRD: National Health Insurance Research Database
NLST: National Lung Cancer Screening Trial
PPV: positive predictive value
t-SNE: t-distributed stochastic neighbor embedding
WHO: World Health Organization
XGBoost: extreme gradient boosting

Edited by R Kukafka; submitted 04.12.20; peer-reviewed by KL Ong, V Montmirail, JA Benítez-Andrades; comments to author 13.02.21; revised version received 03.04.21; accepted 04.05.21; published 03.08.21.

Please cite as:
Yeh MCH, Wang YH, Yang HC, Bai KJ, Wang HH, Li YC(
Artificial Intelligence–Based Prediction of Lung Cancer Risk Using Nonimaging Electronic Medical Records: Deep Learning Approach
J Med Internet Res 2021;23(8):e26256
URL: https://www.jmir.org/2021/8/e26256
doi:10.2196/26256
PMID:34342588

©Hsuan-Chia Yang, Yu-Hsiang Wang, Hsuan-Chia Yang, Kuan-Jen Bai, Hsiao-Han Wang, Yu-Chuan (Jack) Li. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 03.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of

Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Construction of Genealogical Knowledge Graphs From Obituaries: Multitask Neural Network Extraction System

Kai He[1,2,3*], PhD; Lixia Yao[4*], PhD; JiaWei Zhang[1,2,3], MSc; Yufei Li[1,2,3], PhD; Chen Li[1,2,3], PhD

[1]School of Computer Science and Technology, Xi’an Jiaotong University, Xi’an, China
[2]National Engineering Lab for Big Data Analytics, Xi’an Jiaotong University, Xi’an, China
[3]Shanxi Province Key Laboratory of Satellite and Terrestrial Network Technology Research and Development, Xian, China
[4]Department of Health Sciences Research, Mayo Clinic, Rochester, MN, United States
[*]these authors contributed equally

Corresponding Author:
Chen Li, PhD
School of Computer Science and Technology
Xi’an Jiaotong University
Xianning West Road, 27th
Xi’an, 0086 710049
China
Phone: 86 158 0290 2703
Email: cli@xjtu.edu.cn

Abstract

Background: Genealogical information, such as that found in family trees, is imperative for biomedical research such as disease heritability and risk prediction. Researchers have used policyholder and their dependent information in medical claims data and emergency contacts in electronic health records (EHRs) to infer family relationships at a large scale. We have previously demonstrated that online obituaries can be a novel data source for building more complete and accurate family trees.

Objective: Aiming at supplementing EHR data with family relationships for biomedical research, we built an end-to-end information extraction system using a multitask-based artificial neural network model to construct genealogical knowledge graphs (GKGs) from online obituaries. GKGs are enriched family trees with detailed information including age, gender, death and birth dates, and residence.

Methods: Built on a predefined family relationship map consisting of 4 types of entities (eg, people’s name, residence, birth date, and death date) and 71 types of relationships, we curated a corpus containing 1700 online obituaries from the metropolitan area of Minneapolis and St Paul in Minnesota. We also adopted data augmentation technology to generate additional synthetic data to alleviate the issue of data scarcity for rare family relationships. A multitask-based artificial neural network model was then built to simultaneously detect names, extract relationships between them, and assign attributes (eg, birth dates and death dates, residence, age, and gender) to each individual. In the end, we assemble related GKGs into larger ones by identifying people appearing in multiple obituaries.

Results: Our system achieved satisfying precision (94.79%), recall (91.45%), and F-1 measures (93.09%) on 10-fold cross-validation. We also constructed 12,407 GKGs, with the largest one made up of 4 generations and 30 people.

Conclusions: In this work, we discussed the meaning of GKGs for biomedical research, presented a new version of a corpus with a predefined family relationship map and augmented training data, and proposed a multitask deep neural system to construct and assemble GKGs. The results show our system can extract and demonstrate the potential of enriching EHR data for more genetic research. We share the source codes and system with the entire scientific community on GitHub without the corpus for privacy protection.

(J Med Internet Res 2021;23(8):e25670) doi:10.2196/25670

KEYWORDS

genealogical knowledge graph; EHR; information extraction; genealogy; neural network

Introduction

Anthropologists often use oral interviews, historical records, genetic analysis, and other means to obtain genealogical information and draw family trees. When combined with a detailed medical history and social and economic relationships, family trees are considered the x-ray of the family and have been used by clinicians to assess disease risk, suggest treatments, recommend changes in diet and other lifestyle habits, and determine a diagnosis. In the United States, the Medicare Access and CHIP Reauthorization Act of 2015 [1] and Meaningful Use program [2] have incentivized the growing adoption of electronic health records (EHR) with the goal to improve the quality of health care delivery systems. Consequently, a vast amount of EHR data has become available for research purposes in the past decade. However, most EHR systems today do not capture the family relationships between patients by design. Nor do they capture the death information unless patients die in the health care system or the EHR system is linked to external death registries. Constructing family trees for patients becomes an urgent need to unlock the full potential of EHR data in understanding disease and trait heritability, evaluating individuals' health risks, and exploring environmental effects on human health.

Early exploratory works have combined EHR data and family trees for biomedical research. For instance, Mayer et al [3] used twin or multiple relationships to assess the concordance rates for muscular dystrophy and fragile X syndrome in the twin cohort. Schalkwyk et al [4] conducted interviews with family members to build 3-generation family trees with medical chronologies and demonstrated their use in deciding the services required for the psychological well-being of all family members. Wang et al [5] combined diagnosis codes and dependent coverage under medical plans to estimate the heritability and familial environmental patterns of 149 diseases and inferred the genetic and environmental correlations between 29 complex diseases [5]. Similarly, Polubriaginof et al [6] built more than 595,000 family trees from emergency contact information in a large EHR system and then estimated the heritability of 500 traits.

Constructing high-quality large family trees has been challenging. Historically, only famous politicians, philosophers, scientists, religious groups, or royal families were tracked elaborately by genealogists. For such reason, large databases of family trees rarely existed, despite their research value. Recently, a few studies automated family tree collection using innovative informatics approaches. For instance, Mayer and colleagues [3] used shared dates of birth and last names, in addition to home addresses, billing accounts, and keywords of "twin" and "triplet" in unstructured clinical notes to identify a cohort of 19,226 twins or multiples in an extensive health care delivery system. Wang et al [5] inferred 128,989 nuclear families from a large medical claims database covering one-third of the US population based on dependent coverage. Polubriaginof et al [6] used the emergency contact information of 3,550,598 patients from three large EHR systems in New York City to build 595,000 pedigrees. However, these indirect sources, like dependent coverage and emergency contact, have inherent limitations for inferring genealogical information: they do not differentiate biological from nonbiological relationships and they cover only limited types and numbers of family relationships. More specifically, medical insurance in the United States is limited to a beneficiary's spouse and dependents up to age 26 years. Most patients only provided one or two emergency contacts rather than their whole families in their medical records. Missing relationships could be substantial.

Inspired by the work of Tourassi et al [7] and Yoon et al [8], we began to explore online obituaries as a novel data source for the automatic extraction of genealogical information. Obituaries generally cover many more family members with more detailed and accurate descriptions of their family relationships. Nowadays, local newspaper and funeral service companies often publish obituaries on internet, making the cost of obtaining obituaries minimal. In our previous work, we developed and evaluated a new method of name entity recognition (NER) for extracting family members' names and relation classifications (RCs) for classing the pairs of names among family members mentioned in online obituaries [9]. In this work, we advanced our previous work in the following 5 aspects: (1) for the NER task, we processed more entity types, including people's name, age, residence, and dates of birth and death; (2) for RC, we also matched residence entity and related people (in the previous work, we only extracted the family relationships among people entity); (3) we parsed two kinds of special language patterns in obituaries, last name distributive and name with parentheses; (4) all the triplets of family relationships were integrated to build the enriched family trees with additional rule-based inference; and (5) in terms of training data, we normalized the family relationships (see details in Data section).

Traditionally, NER and RC were considered two separate tasks for information extraction. NER sought to extract named entities mentioned in unstructured text into predefined categories, whereas RC classified the relations between those extracted entity mentions. Researchers built natural language processing (NLP) pipelines with multiple modules to accomplish specific tasks. However, such modular separation suffered from 3 major issues leading to suboptimal results: (1) errors from the NER propagated to RC, (2) it was computationally redundant and time-consuming as the system had to pair up every two named entities to classify their relations, and (3) the pipeline model could not take full advantage of the knowledge inhabitant in the relationships of 2 or more named entities. For instance, if the system detected a *live in* relationship between two named entities in obituaries, the first entity is likely to be a person's name and the second entity is likely to be a location.

Thus, we look at multitask models that can simultaneously handle multiple related tasks and optimize their learning abilities by sharing the knowledge learned in all or some of the tasks [10]. In 2008, Collobert [11] introduced a single neural network architecture that solved NLP tasks such as part-of-speech tagging, chunking, named entity recognition, semantic role identification, and semantically similar word grouping using one language model. Recently, there are 3 prevailing solutions for multitask NLP models. The most popular solution establishes a common neural network presentation space for all tasks followed by task-specific classifiers [12,13]. The second proposes novel neural network architecture for multiple tasks. Sun et al [14] used graph convolutional networks to enhance interaction between entity and relation. Bhatia et al [15] proposed a hierarchical encoder-decoder NER model and adopt a shared encoder followed by separate decoders for NER and RC tasks. The third focuses on customized tagging schema. For instance, Zheng et al [16] proposed a novel tagging schema for long short-term memory (LSTM) models that simultaneously identified named entities and extracted relationships in a corpus of New York Times news. In addition, Dixit et al [17] introduced a span-level solution to handle NER and RC together. Zheng et al [18] introduced a hierarchical solution that combines an encoder-decoder LSTM module for NER with a convolution neural network for RC.

In this work, we first updated our annotated corpus by defining a family relationship map to normalize various family relations (see details in Data section). We also used data augmentation technology to generate more synthetic data (sentences), in order to address the imbalanced training data issue and boost the performance on rare classes [19]. After that, we proposed an end-to-end information extraction system based on a multitasking solution. The end-to-end system included a knowledge inference layer for gender inference based on name and relationship mentioning. In the end, we constructed family trees centered on the deceased. These family trees contained many family members with detailed information, including age, gender, death date, birth date, and residence. We named such enriched family trees genealogical knowledge graphs (GKGs). These GKGs could be linked to external EHR data in Minnesota by personally identifiable information (PII), in a similar way as Sauver et al [20] did. We empirically estimated the upper bound of the mapping precision could be around 80% to 90%. It would significantly enhance the power of EHR data to study disease and trait heritability, evaluate an individual's health risks, and explore environmental effects on the human health.

Methods

Data

We collected 12,407 obituaries published from October 2008 to September 2018 from 3 funeral services websites and 1 local newspaper in the Twin Cities area, metropolitan Minneapolis–Saint Paul. Our data sources were limited to openly available obituaries. Considering the PII embedded in online obituaries, we decided to take a cautious and conservative position in our work by marking up the last name of any real people with the symbol XX (see more details on privacy protection in the Discussion section). After data cleaning, we randomly sampled 1700 obituaries for annotation. We developed the annotation guideline and trained 3 annotators to annotate each of the 1700 obituaries independently. The interannotator agreement measured by F-1 was 82.80% [9]. Table 1 shows the summary statistics of the annotated corpus. There were two unique language patterns in obituaries, namely last name distributive and name with parentheses (see Table 2 for examples). These patterns might be due to the word limitation when the family paid for publishing an obituary in printed newspapers. They required special treatment, as described in the next session of end-to-end system.

Table 1. Summary statistics of the annotated corpus.[a]

Corpus	Count	Deceased person	Count	Special language patterns	Count
Sentences	28,317	Full name	1551	Last name distributive	4954
Names	27,108	Age	1379	Name with parentheses	7504
Family relationship	25,557	Death date	1557	Spouse's name	5993
Residence	7161	Birth date	1368	Previous last name	1511
Name-residence pair	7954	—[b]	—	—	—

[a] All counts are the number of occurrences except for the full name of the deceased. Considering all obituaries have structured metadata giving the full names of the deceased more precisely, we only annotate and extract the first-time mention of a full name of the deceased in an obituary. Spouse's name and previous last name are 2 categories in the content name with parentheses.

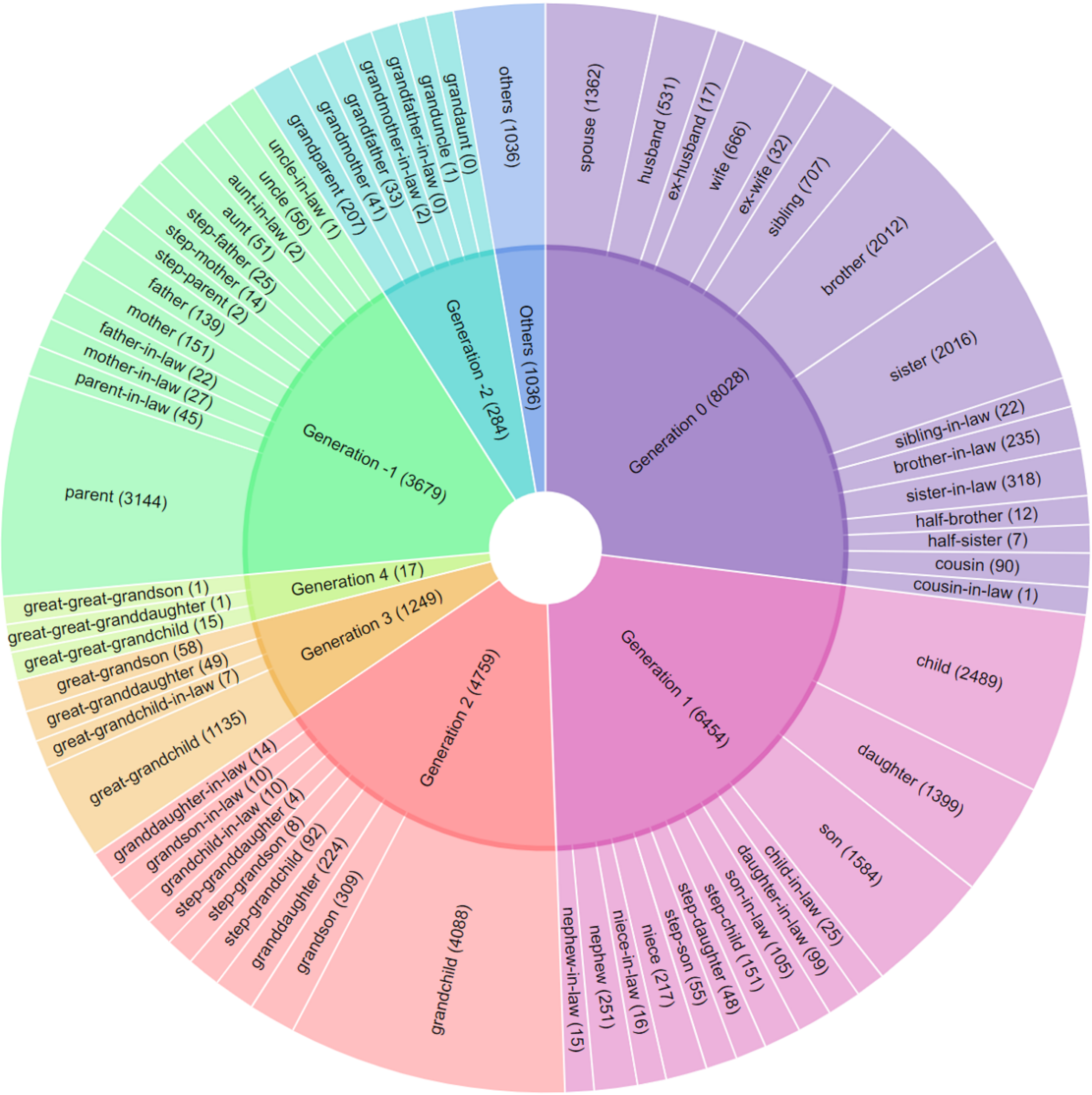
[b] Not applicable.

Table 2. Examples of unique language patterns in obituaries.

Language pattern	Example	Explanation
Last name distributive	He is survived by grandsons Addison and Owen XX.	XX is also the last name for Addison.
Names with parentheses		
Previous last name	Anne was born March 20, 1952, to William and Isabel (Starr) XX.	Starr is the maiden name for Isabel XX.
Spouse's name	Survived by her sons, Dale (Mary) and Bruce (Diana).	Dale's wife is Mary, and Bruce's wife is Diana.

In this work, we made two improvements in the corpus annotations. First, we created a family relationship map that normalized various family relationship mentions to 71 family relationship groups. For example, there were many mentions of "born to (name)," "daughter of (name)," and "son of (name)" in obituaries, which were equivalent to express the parent of the deceased. We grouped them into the "parent" relation. Similarly, we treated "married to" the same way as the "spouse (of)" relation. Figure 1 shows the family relationship map, consisting of 8 generations and 71 normalized family relationships. The numbers in parentheses were the number of occurrences of a specific family relationship in our corpus.

Figure 1. Family relationship map in the obituary corpus.



It was observed that some family relationships, such as granduncle, uncle-in-law, and half-sister had small numbers of cases that was not sufficient to train a high-performance neural network model. Therefore, we used data augmentation technology [19] to expand the corpus and alleviate the imbalanced data issue. We first introduced w_i, the weight of relation i:

Where c_i stood for the count of annotated sentences with relation i, and n was 71, the count of all family relationship groups defined in the family relationship map. For each family relationship i, the number of training sentences to be generated, g_i, was computed as follows:

Where N was the total number of all human annotated sentences and was the user-defined ratio for data augmentation. Essentially we generated more synthetic sentences to ensure each family relationship had no less than 200 examples, with the constraint that the count ratios of all family relationships remain as close as possible to those in the original training data.

After deciding g_i, 2 steps were performed to generate extra sentences. First, we randomly chose g_i sentences from the raw corpus and replaced one of the raw family relationship tokens in these sentences with relationship word *i*. Second, we randomly chose one of the following operations introduced by Wei and Zou [19] to generate the final augmented sentences:

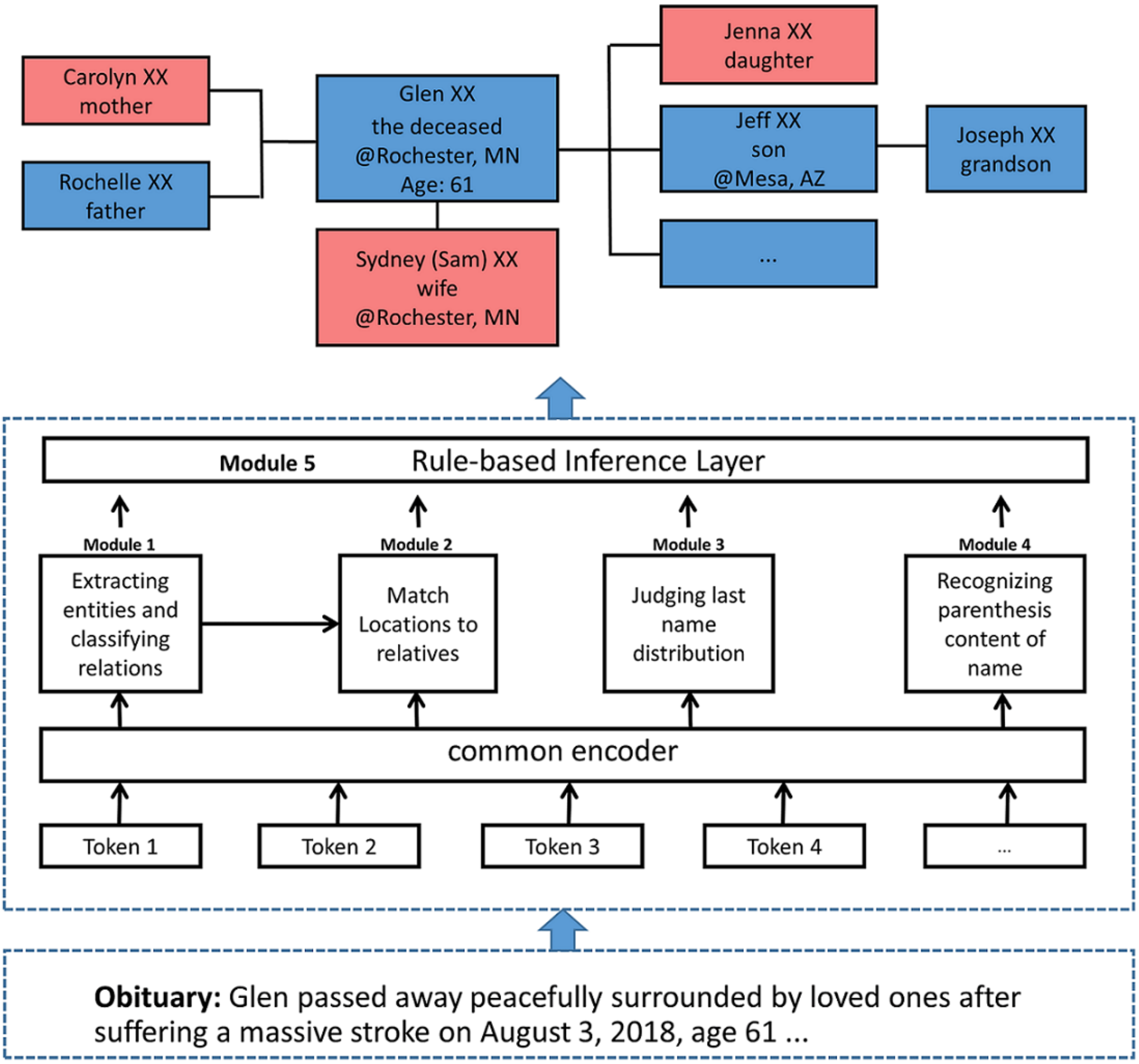
- Synonym replacement: randomly replace *n* non–stop words with their synonyms in the sentence
- Random insertion: randomly insert a word's synonym before or after the chosen non–stop word in the sentence
- Random swap: randomly swap 2 non–stop words in the sentence
- Random deletion: randomly remove a non–stop word in the sentence

The 4 entity types of interest in this work, name, residence, birth date and death date, are exempt from the changes. It should also be noted that the generated sentences could not be guaranteed to be grammatically and semantically correct. However, for neural network models, such sentences, when created with appropriate α, were demonstrated to improve models' generalizability as noisy training data.

End-to-End System

Figure 2 illustrates our end-to-end system. It took a list of segmented sentences in an obituary as the input and generated a GKG centered around a deceased person. Its core was a multitask system that combined common parameter sharing across different modules and custom tagging schemes. The multitask solution promised better performance, as it used more supervision information and understood data from different views [21]. The 4 modules were (1) named entity recognition and relation classification through a joint training model and customized tagging scheme, (2) matching locations to people's name, (3) a parser for resolving last name distributive, and (4) a parser for resolving names with parentheses. These 4 modules shared the same model parameters, as they were trained jointly using one common weighted loss function. Among these modules, module 2 needed the extracted names and locations from module 1 as inputs. Module 5 was added as an independent rule-based layer for gender inference and age, date of death, and birth inference. Eventually, the results of these modules were combined to construct the GKGs.

Figure 2. End-to-end extraction system to parse obituaries and generate genealogical knowledge graphs.

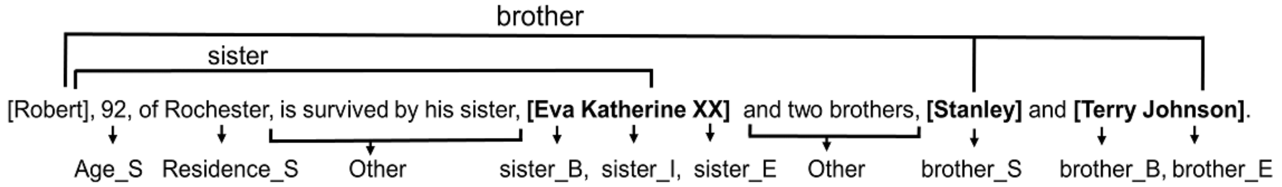

Carolyn XX
mother
Rochelle XX
father
Glen XX
the deceased
@Rochester, MN
Age: 61
Sydney (Sam) XX
wife
@Rochester, MN
Jenna XX
daughter
Jeff XX
son
@Mesa, AZ
Joseph XX
grandson
...
Module 5
Rule-based Inference Layer
Module 1
Extracting entities and classifying relations
Module 2
Match Locations to relatives
Module 3
Judging last name distribution
Module 4
Recognizing parenthesis content of name
common encoder
Token 1
Token 2
Token 3
Token 4
...
Obituary: Glen passed away peacefully surrounded by loved ones after suffering a massive stroke on August 3, 2018, age 61 ...

Module 1: Joint NER and RC

This module aimed to extract family members' names, relationships, and additional attributes of people (residence, age, death date, birth date). Gender was usually not explicitly mentioned in the obituaries, so we inferred the gender in module 5. We adopted a customized tagging scheme (shown in Figure 3) when annotating the training data. Each tag consisted of 2 parts. The first part indicated the type of an entity, and the second part illustrated the position of the word in the entity. As shown in Figure 3, "sister_B," "sister_I," "sister_E," and "Age_S" indicated the beginning, the inside, and the end of a sister entity and a single-word entity of age, respectively. In the system, the deceased was the default baseline entity for all family relationship triplets. In the sentence shown in Figure 3, for example, "Robert" was the name of the deceased person (we knew it from the obituary metadata and the context of the entire obituary). After annotation, we obtained three triplets (Robert, sister, Eva Katherine XX), (Robert, brother, Stanley), and (Robert, brother, Terry XX). The calculating process was as follows:

For each input token x_i, we used BERT [22] as a common encoder to obtain each hidden representation h_i^{common}. Then h_i^{common} were sent into one LSTM classifier to obtain each tag $T_{1i} \in T_1$, where T_1 was the result set of module 1, and w_1 and b_1 were parameters for training.

Figure 3. Tagging scheme for simultaneously extracting entities and kinship. S: single; B: begin; I: inside; E: end.


brother
sister
[Robert], 92, of Rochester, is survived by his sister, [Eva Katherine XX] and two brothers, [Stanley] and [Terry Johnson].
Age_S
Residence_S
Other
sister_B, sister_I, sister_E
Other
brother_S
brother_B, brother_E

Module 2: Matching Locations to People

After identifying the residence entities (eg, Rochester in Figure 3), we need to match them with specific people. To do so, we used 3 inputs, all extracted names $T_1^{\text{name}} \in T_1$, all extracted residences $T_1^{\text{residence}} \in T_1$, and common representation h_i^{common}. This module followed by a co-reference solution [23]. We defined the process as follows:

where v_{1i}^{name} denoted the vector of one name entity $t_{1i}^{\text{name}} \in T_1^{name}$, $v_{1j}^{residence}$ denoted the vector of one residence entity $t_{1j}^{\text{residence}} \in T_1^{residence}$, [] denoted concatenate, * denoted dot product, w_2 and b_2 were one linear layer parameters for training, $T_{2k} \in T_2$ was the matching result for each pair of name and residence, and T_2 was the final selected pair of name and residence.

Module 3: Judging Last Name Distributive

We identified 2 special language patterns in obituaries, last name distributive and names with parentheses, as shown in Table 2. Resolving these language patterns was helpful for extracting and constructing high-quality GKGs. The task of module 3 was to decide for each token in an input sentence if the last name distributive existed by assigning each token with a binary tag of yes OR no. When we cotrained module 3 with other modules, these tags would concatenate with other modules' tags for joint training. In the sentence in Figure 3, for example, "Stanley" and "Terry" shared the same last name of "Johnson." Therefore, in module 3, "Stanley" was assigned a label "brother_S_yes" and "Terry Johnson" was given 2 tags "brother_B_yes" and "brother_E_yes." This way, the module would extract 2 full names, *Stanley Johnson* and *Terry Johnson*, instead of *Stanley* and *Terry Johnson*. The detailed computing process was as follows:

where w_4 and b_4 were parameters for training, $T_{4i} \in T_4$ was the result for each name, and T_4 was the result set of module 3.

Module 4: Recognizing Names With Parentheses

Module 4 was a 3-class classifier to determine whether there was a parenthesis in a name, and if so, whether it referred to a previous last name or the name of spouse. The computing process was the same as module 3, which took the input of h_i^{common} and output the tags of 3 classes ("no parenthesis," "previous last name," and "spouse's name").

Module 5: Rule-Based Inference Layer

This module aimed to infer age, death date, and birth date for the deceased and gender for both the deceased and their family members. First, if an obituary mentioned any 2 attributes out of age, birth date, and death date for the deceased, we calculated the third one. Second, we used both family relationship keyword and name to infer gender. If a family relationship keyword (eg, son, daughter, nephew) suggested gender, we would add the gender tag accordingly. Otherwise, when the family relationship keyword (eg, spouse and parent) did not tell the gender, we used an external human name knowledge base to match the most likely gender with names. For instance, "Tom" and "Emily" indicated male and female, separately.

After constructing the GKGs from each obituary by modules 1 to 5, we assembled the extracted GKGs into bigger ones by matching PII, including people's names, residence, birth date, death date, and family relationship.

Joint Training Loss

We minimized the negative log likelihood loss of the generated tags for the first 4 modules (module 5 is a rule-based inference layer that did not require training). For module *k* (*k*=1, 2, 3, 4), the loss function was defined as follows:

Where *B* was the batch size, l_s was the length of input sentence sentences, y^s_i and p^s_i were the true tag and the normalized probability of the predicted tag for an input token *I*, and was a hyperparameter. *P(O)* was the indicator function that determined which part of equation 10 was used to calculate the loss. If the current tag was not "O" (other), the hyperparameter would decide the weight of the loss function. It was defined as follows:

In the end, we combined all four loss functions L_1, L_2, L_3 and L_4 together, using different weighting parameters λ_k into the final loss function, which was optimized for the entire training as follows:

Evaluation Metrics

We performed 10-fold cross-validation by randomly selecting 10% of the annotated data for validation and the remaining for training. It is worth noting that the augmented data were only used for training models. Extracted GKGs consists of outputs from modules 1 to 5. They were measured by averaged performance of all modules except module 5 due to this rule-based inference module lacking a gold standard. For modules 1 to 4, we used precision, recall, and F-1 measure for evaluation, which were computed as follows:

In module 1, the outputs were entity mentions with extra entity and relation types. We defined an extracted mention as true positive instances only if the mention's boundary, entity type, and relation tags were exactly matched with the gold annotation. The instances of false positive were predicted mentions that do not precisely match with gold annotation boundaries, entity, or relation types. False negative instances were those existing in the gold annotation but not recognized by the model.

In module 2, true positive instances were defined as pairs of name and location that matched exactly. If either name or location was wrong, the pair would be considered a false positive. False negative referred to the name-location pairs missed by our system.

Module 3 and module 4 were formulated as generic classification tasks, so we used common definitions of false negative, false positive, and true positive. For all modules, evaluation metrics were precision, recall, and F-1 measure.

Results

Table 3 illustrates the performance of modules 1 to 4 with ablation experiments in terms of macroaveraged and microaveraged precision, recall, and F-1 measure (without data augmentation). A macroaverage is the arithmetic average of the computed metrics for all classes and a microaverage sums up all true positive, true negative, false positive, and false negative instances before computing the final precision, recall, or F-1 measure for all classes. Macroaveraged metrics are often used for evaluation, particularly when there are extremely imbalanced classes, as no single class should largely dominate the results.

As shown in Table 3, we can see that module 1, which constructed the nodes and edges of the GKGs, achieved macroaveraged precision, recall, and F-1 measure of 83.85%, 83.05%, and 83.44%, respectively (see the third row of the macroaveraged performance). If we did not consider the effects of imbalanced data, the microaveraged precision, recall, and F-1 measure were even better, reaching 95.42%, 93.52%, and 94.46%, respectively (see the third row of the microaveraged performance). The macroaveraged results were much worse than the microaveraged results because the dominating classes had above-average performance. The minority classes, although having below-average performance, did not affect the microaveraged results much due to their small count numbers. For example, uncle-in-law, granduncle, and cousin-in-law had just 1 case in our corpus (see Figure 1). These relations affected the macroaveraged metrics more negatively. The performances for modules 2, 3, and 4 had similar patterns.

We also observed the benefits of multitask models through ablation experiments. Extra information gained from modules 3 and 4 seemed to improve module 1 in both macroaveraged precision, recall, and F-1 measure (2.17%, 3.12%, 2.64%, respectively) and microaveraged precision, recall, and F-1 measure (1.27%, 1.12%, and 1.19%, respectively). Modules 1 and 3 improved the performance of module 4 by 2.76%, 1.32%, 2.08% for macroaveraged precision, recall, and F-1 measure, respectively, and 2.51%, 1.7%, 2.13% for microaveraged precision, recall, and F-1 measure, respectively. Similarly, modules 1 and 4 helped to improve the macro/micro precision, recall, and F-1 of module 3 by 2.74%, 1.00%, 1.88%, respectively. And modules 1, 3, and 4 improved module 2 by 1.10%, 5.17%, and 3.49% in macro/micro averaged precision, recall, and F-1 measure.

It should be noticed that module 2 seemed not helpful in improving the overall performance of each module. For module 1, the macroaveraged and microaveraged F-1 measure dropped by 1.41% (compare the first and third row of the macroaverage section of Table 3) and 1.03% (compare the third and third row of the microaverage section) after introducing module 2 into the end-to-end system . Other modules had similar effects when included in module 2. This phenomenon was named negative transfer. It meant that although module 2 significantly benefited (F-1 measure raised from 75.08% to 78.57%), other modules were negatively affected. Liu et al [24] and Wang et al [25] also observed and discussed similar negative transfer effects. We talk about negative transfer further in the Discussion section.

In our system, the solution for avoiding the negative transfer was that module 1, 3, 4 would be cotrained and module 2 would be separated from the whole system for training. In such a way, each module could benefit the most from the joint training method.

Table 3. Model performance of each module with ablation experiments.

Module and ablation test	Macroaveraged performance			Microaveraged performance		
	P[a] (%)	R[b] (%)	F1[c] (%)	P (%)	R (%)	F1(%)
Module 1						
Baseline	81.68	79.93	80.80	94.15	92.40	93.27
Joint training (module 2, 3, & 4) + negative transfer	82.07	81.99	82.03	94.08	92.79	93.43
Joint training (module 3 & 4)	83.85	83.05	83.44	95.42	93.52	94.46
Module 2						
Baseline	83.17	68.43	75.08	—[d]	—	—
Joint training (module 1, 3, & 4)	84.27	73.60	78.57	—	—	—
Module 3						
Baseline	89.64	92.01	90.81	—	—	—
Joint training (module 1, 2, & 4) + negative transfer	91.48	91.12	91.30	—	—	—
Joint training (module 1 & 4)	92.38	93.01	92.69	—	—	—
Module 4						
Baseline	90.65	94.74	92.64	90.96	95.21	93.03
Joint training (module 1, 2, & 3) + negative transfer	92.34	95.76	94.02	92.37	96.31	94.30
Joint training (module 1 & 3)	93.41	96.06	94.72	93.47	96.91	95.16

[a] P: precision.

[b] R: recall.

[c] F1: F-1 measure.

[d] The microaveraged and macroaveraged performances are the same for module 2 and module 3 because they are both binary classification tasks. All results shown are from the curated corpus without data augmentation.

We also adopted data augmentation technology to expand our corpus, aiming to improve the relation extraction performance for family relations (module 1) with too few training examples. By synonym replacement, random insertion, random swap, and random deletion, we augmented the training data to ensure every relation had no less than 200 training examples. However, the automated data augmentation method introduced new noise. We tested a different augmentation ratio (α) to find the best balance. As shown in Figure 4, when the augmentation ratio was set to 40%, the extra synthetic data in training benefited our model most. It was worth noting that the augmentation data were only used in training for module 1, and we still evaluated our system with real, nonsynthetic test data. Figure 4 shown that the best macroaveraged and microaveraged F-1 measures achieved 89.14% and 95.55%, respectively, for module 1. With augmented module 1, our whole system achieved the best macroaveraged performances, 92.59% (precision), 90.05% (recall), and 91.30% (F-1 measure), and the best microaveraged metrics were 94.79% (precision), 91.45% (recall), and 93.09% (F-1 measure). These results confirmed that data augmentation technology can alleviate the problem of imbalanced data.

After extracting GKGs from all obituaries, we assembled them into bigger ones by matching available PII, including name, gender, age, residence, and birth date. Considering obituaries usually provide detailed PII for the deceased but not for their family members and relatives, we did fuzzy matching for the relatives. That is, if the mentioning of 2 people in 2 different obituaries are likely to refer to the same person based on 1 or more shared piece of PII, we would assemble 2 GKGs into 1. In the end, we had 319 GKGs assembled into 149 bigger GKGs after processing all 12,407 downloaded obituaries. Among those 319 obituaries, 22.3% (71/319) had 1 shared PII item, 8.5% (27/319) had 2, and 69.3% (221/319) had more than 2. We manually evaluated those 149 assembled GKGs and confirmed that 71.8% (107/149) were correct, 12.1% (18/149) were wrong, and 16.1% (24/149) were uncertain. We acknowledge that this rule-based matching method is limitedly useful for the selected geographic location of the Twin Cities area in Minnesota. It might be more error prone to apply to the entire country or other densely populated areas with high population mobility. So we did not include the assembly function in the end-to-end system but kept it as an additional resource for cautious users.

Figure 4. Comparing the F-1 measures of raw corpus and augmented corpus.

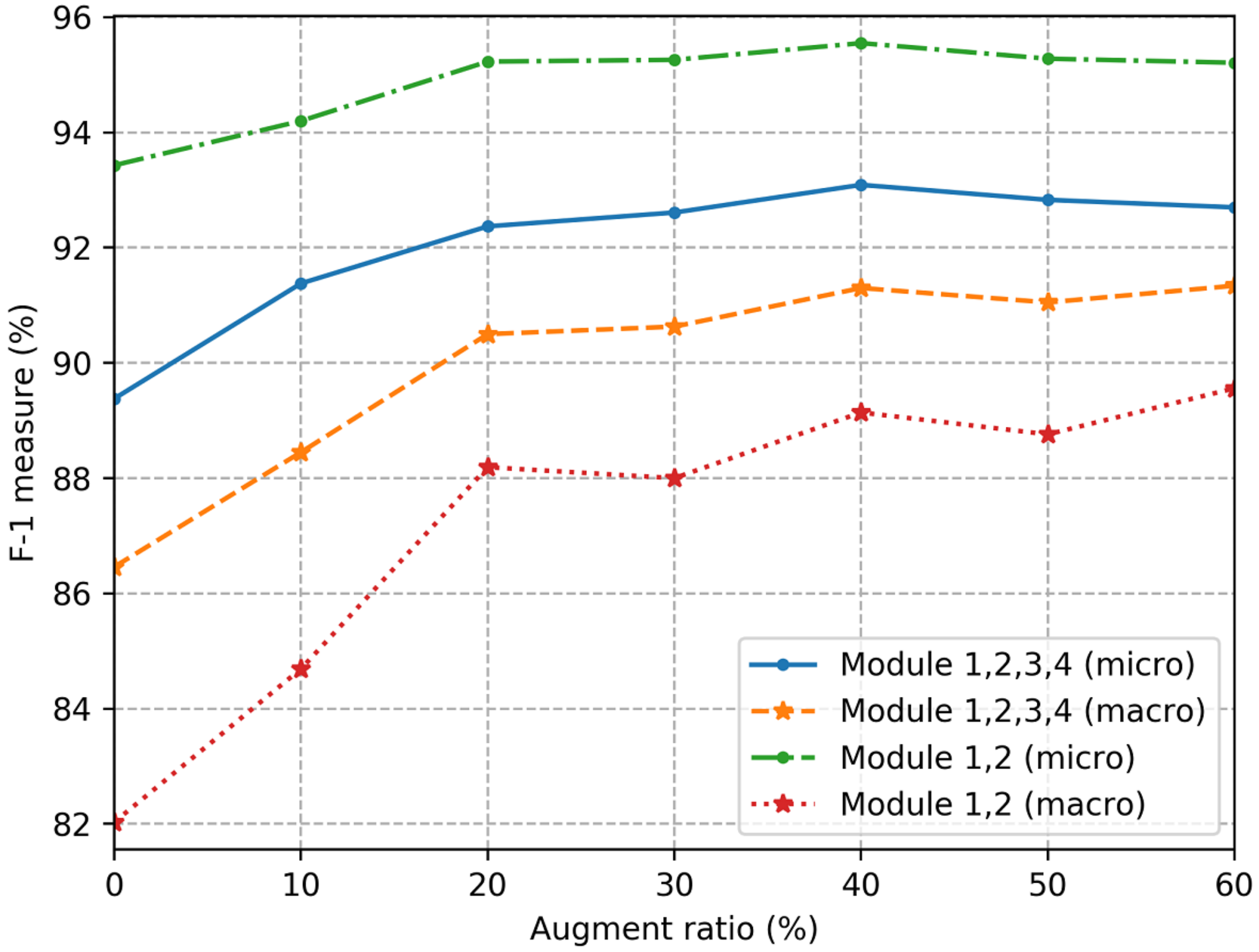


Figure 5 shows one example of assembled GKG from 3 obituaries. It contained 4 generations and 30 people. Figure 6 is the corresponding gold standard result conducted from manual validation. It can be seen that the assembled GKG missed the state name Minnesota for Dorothy and Patrick's residences and one family member, Joe, who was Lynne's husband (missing parts are shown in dashed boxes). In the original obituary, the sentence mentioning Lynne and Joe's relation was "...he proposed...they began 54 years of happy life," and our system failed to capture this subtle language. The successful assembly of multiple obituaries also demonstrated the feasibility of linking family relations extracted from obituaries to EHRs to support genetic research like linkage analysis and disease risk prediction. Meanwhile, it should be noticed that even though obituaries inherently contained rich genealogical information and the system extracts the GKGs with high accuracy, the GKGs should not always be equated to pedigrees used by genealogists. Although it is common to declare blood or nonblood relationships in obituaries due to data specialty (detail analysis for the slippery slope of genealogy issue shown in the Discussion section), we cannot guarantee people always declare the difference of blood or nonblood and always list all of their family members for various reasons.

Figure 5. An example of an assembled genealogical knowledge graph. We removed last names for privacy protection. The symbol ? means we are not sure which children nodes belong to which parent nodes.

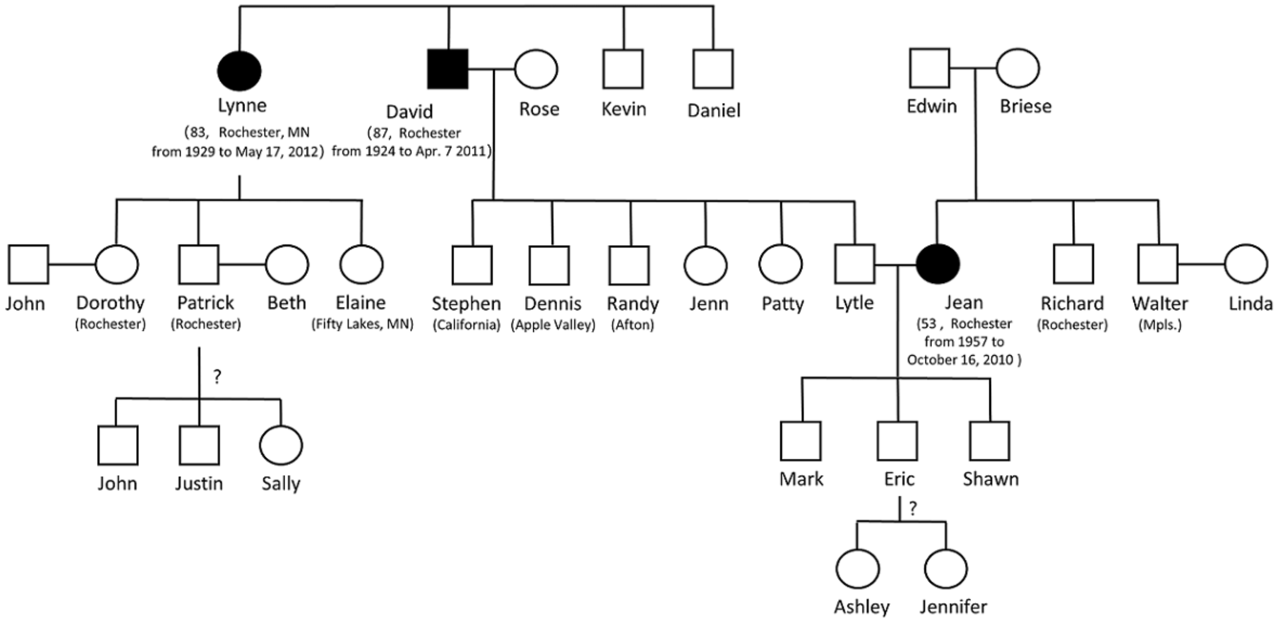
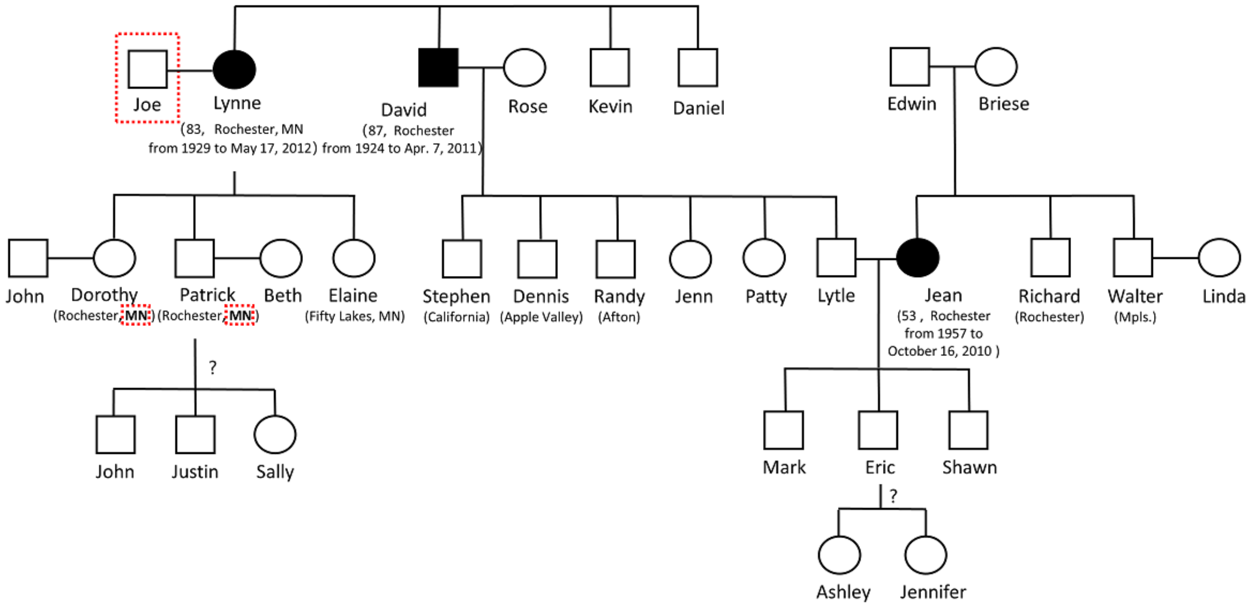

Lynne
(83, Rochester, MN
from 1929 to May 17, 2012)
David
(87, Rochester
from 1924 to Apr. 7 2011)
Rose
Kevin
Daniel
Edwin
Briese
John
Dorothy
(Rochester)
Patrick
(Rochester)
Beth
Elaine
(Fifty Lakes, MN)
Stephen
(California)
Dennis
(Apple Valley)
Randy
(Afton)
Jenn
Patty
Lytle
Jean
(53 , Rochester
from 1957 to
October 16, 2010)
Richard
(Rochester)
Walter
(Mpls.)
Linda
?
John
Justin
Sally
Mark
Eric
Shawn
?
Ashley
Jennifer

Figure 6. The gold standard family tree constructed from manual curation corresponding to Figure 5.


Joe
Lynne
(83, Rochester, MN
from 1929 to May 17, 2012)
David
(87, Rochester
from 1924 to Apr. 7, 2011)
Rose
Kevin
Daniel
Edwin
Briese
John
Dorothy
(Rochester, MN)
Patrick
(Rochester, MN)
Beth
Elaine
(Fifty Lakes, MN)
Stephen
(California)
Dennis
(Apple Valley)
Randy
(Afton)
Jenn
Patty
Lytle
Jean
(53 , Rochester
from 1957 to
October 16, 2010)
Richard
(Rochester)
Walter
(Mpls.)
Linda
?
John
Justin
Sally
Mark
Eric
Shawn
?
Ashley
Jennifer

Discussion

Principal Findings

In this work, we proposed an end-to-end system to construct GKGs from online obituaries, aiming at supplementing EHR data for genetic research. This system achieves microaveraged precision of 94.79%, recall of 91.45%, and F-1 measure of 93.09% after data augmentation technology. The work exploits the large availability of obituaries on the internet, which are consistent with the vital records and census records and more reliable and comprehensive than dependent information from medical insurance and emergency contact in EHR systems [5,6]. We demonstrate an efficient system to automatically build large GKGs from 10 years' online obituaries in the Twin City area, Minnesota. Furthermore, by identifying individuals, we explore integrating related GKGs into bigger GKGs and manually validating the integrated results. The results show the feasibility of identifying individuals by extracted information, including residence, age, gender, birth, and death dates. We compute similarities between GKGs to further merge them into more complete GKGs. In the future, the similarity computing techniques could assist mapping the GKGs to the EHRs.

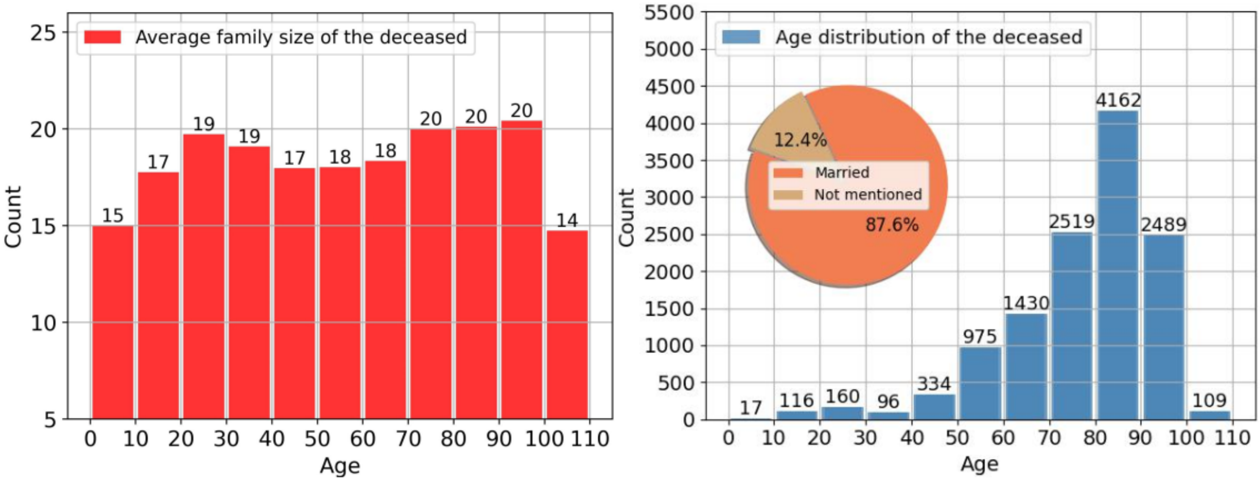
In this work, we use publicly available obituaries. The Association of Internet Researchers, in partnership with their Ethical Working Committee, formulated general principals to guide online research [26]. While this document presents the overarching ethical considerations relevant to social media–based research, a comprehensive determination of ethical principles and best practices has yet to be developed. Furthermore, debate continues as to whether some forms of social media–based research, namely analysis of existing textual archives (strictly speaking, online obituaries are not social media, but they have similar characteristics as a data source for biomedical research), fall within the parameters of human subject research or constitute an alternative form of humanistic inquiry [27]. Considering the PII embedded in online obituaries, we decided to take a cautious and conservative position in our work by marking up the last name of any real people mentioned in the paper.

As a novel data source, obituaries are informative for constructing family trees. It is hard to obtain such rich genealogical information from other data sources, but there are caveats to their use as genealogical data. First, semantic ambiguity occurs in obituaries as it occurs in many other types of human writing. For example, it is not uncommon to see statements like "...survived by two sons, Marshal and Paul XX and daughter Daisy, and four grandchildren Denny, Gary, Cecil, and Alina." In this case, it is impossible to tell the exact parents for each of the 4 grandchildren Denny, Gary, Cecil, and Alina. All we know is that their parents are Marshal XX, Paul XX, and Daisy. Additional data sources like birth certificate registries can be helpful in this case.

A second point worth discussing is the slippery slope of genealogy. Compared with medical insurance and emergency contact information [5,6], a statement of nonblood relationship is more common in obituary data due to their specificity. As shown in Figure 1, for child relationship the ratio for nonblood versus blood is 483:5472 (there are 25 mentions of child-in-law, 99 of daughter-in-law, 105 of son-in-law, 151 of stepchild, 48 of stepdaughter, and 55 of stepson compared with 2489 cases of child, 1399 of daughter, and 1584 of son). A similar ratio can be observed in nonblood parent relationship. This advantage could be helpful for alleviating the problem of the slippery slope of genealogy. However, it is still worth mentioning that not all people make such distinctions in obituaries.

In addition, Figure 7 displayed the related statistics aimed at showing potential data bias. We plotted the distribution of age (at death), average number of mentioned family members, and marital status of the deceased for all GKGs extracted from 12,407 downloaded obituaries. As shown in Figure 7, the age distribution of the deceased is consistent with public health data (73.9% of the deceased died at the ages of 70 to 100 years). The average numbers of mentioned family members seem similar for different age groups; only those died in the 0 to 10 and 100 to 110 age groups had relatively smaller family size (≤15); 87.6% of the GKGs indicated that the deceased was married at least once. We did not interpret the results too deeply because we did not have a good understanding of the sample bias. Meanwhile, it was noticed that people who had complete and/or affluent families tended to publish obituaries. Although these data biases would not affect the performance of our extraction system, the fact that extracted GKGs may be biased should be considered when researchers are using them in other research.

Figure 7. Left: distribution of average numbers of mentioned family members. Right: age and marital status of the deceased person in 12,407 extracted genealogical knowledge graphs.



Technically, the data used in the research are very imbalanced, in which 14 rare relationships have fewer than 10 instances. We adopted the augmentation technology to enhance system performance. For example, in the relationships half-sister, grandchild-in-law, and grandson-in-law, their F-1 measures increased from 20.0%, 30.0%, and 35.71% to 66.67%, 50.0%, and 71.43%, respectively. Next step, we plan to experiment with additional few-shot (extremely imbalanced)–based information extraction and mate learning to improve the system [28,29].

In our end-to-end solution, the performance of module 2 was obviously inferior to the other modules. Besides the error propagation problem (module 2 need the results from module 1), the task of module 2 was a semantic matching resolution problem, which is still challenging in the NLP community. In addition, we currently have curated an obituary corpus in English to train the neural network models. To expand to other languages, a new corpus in those specific languages and new gender inference rules would need to be curated. There is some cross-language transfer research in the NLP community which suggests neural models trained on an English corpus can help to build NLP models in other languages by reducing training data and training time. Sometimes such transfers even provide more robust models with better performance [30,31].

In our end-to-end solution, module 2 currently is the bottleneck. This module suffered significantly from negative transfer. Generally speaking, when a task or domain was joined with data of no relatedness or similarity, the added data would become noise rather than useful information. It remains challenging to quantitatively measure the relatedness or similarity among different tasks or domains [32]. Therefore, most transfer learning solutions rely on empirical methods and do not account for negative transfer effects. In this work, we considered module 2, which matched locations to people, as strongly related to other modules that extracted locations or people and paired them. Unfortunately, the experiment results showed negative transfer still occurred. One possible explanation was about the different natures of tasks in modules 1, 2, 3, and 4. Module 2 was a classification task with 2 entity mentions as the input and a class tag as the output. All other modules were sequence tagging tasks, where the whole sentence was the input and tags for all tokens of an input sentence were output. Another possible reason was that the task of module 2 was much more challenging than the others. Modules 1, 3, and 4 all had a higher than 90% microaveraged F-1 measure when we tested them individually, while module 2 had a 75.08% microaveraged F-1 measure. In addition, module 2 needed inputs from module 1. The errors of module 1 would propagate to module 2. How to improve module 2 and alleviate its negative transfer and error propagation is what we plan to focus on methodologically in the future.

Besides the performance benefits shown in the Result section, the multitask solution is also faster to train. We use a single V100 GPU in this study. For the traditional pipeline model, one round 10-fold cross-validation experiment costs about 240 hours in total. However, the multitask model with all 4 modules together takes only 150 hours. For module 1, the training process took about 70 epochs to achieve an F-1 measure of 80% when being trained independently. The multitask method takes less than 5 epochs to achieve the same level of F-1 measure.

Limitations

The first limitation of our work is the existing potential data bias. Our data are collected from online obituary websites. In such conditions, people who had intact and/or affluent families tended to publish obituaries. The second limitation is that our system is mainly for English obituaries. Modules 2 and 3 are designed for 2 English writing patterns.

Conclusions

GKGs have great potential to enhance many medical research fields, especially combined with EHR data. We believe a high-quality, large-scale genealogical information database will have significant research meaning. In this work, we presented a new corpus with a predefined family relationship map and augmented training data and proposed a multitask deep neural system to construct and assemble GKGs. With the data augmentation technology, the system achieved microaveraged precision, recall, and F-1 measure of 94.79%, 91.45%, and 93.09%, respectively, and macroaveraged precision, recall, and F-1 measure of 92.59%, 90.05%, 91.30%, respectively. Based on such promising results, we developed PII-matching rules to assemble large GKGs, demonstrating the potential of linking GKGs to EHRs. The system is capable of generating a large number of GKGs to support related research, like genetic research, linkage analysis, and disease risk prediction. We share the source codes and system with the entire scientific community on GitHub, without the corpus for privacy protection [33].

In the future, we will improve the performance of our system further and match GKGs with more medical information, like EHR databases. With the massive obituary data freely available on the internet or other textual data that contain genealogical information, our ultimate goal is to accelerate large-scale disease heritability research and clinical genetics research.

Acknowledgments

This work has been supported by grant 2018YFC0910404 from the National Key Research and Development Program of China, grant 61772409 from the National Natural Science Foundation of China; grant 61721002 from the Innovative Research Group of the National Natural Science Foundation of China, and grant IRT_17R86 from the Innovation Research Team of the Ministry of Education, Project of China Knowledge Centre for Engineering Science and Technology.

Conflicts of Interest

None declared.

References

1. Casalino LP. The Medicare Access And CHIP Reauthorization Act and The corporate transformation of American medicine. Health Aff (Millwood) 2017 May 01;36(5):865-869. [doi: 10.1377/hlthaff.2016.1536] [Medline: 28461353]
2. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. N Engl J Med 2010 Aug 5;363(6):501-504. [doi: 10.1056/NEJMp1006114] [Medline: 20647183]
3. Mayer J, Kitchner T, Ye Z, Zhou Z, He M, Schrodi SJ, et al. Use of an electronic medical record to create the marshfield clinic twin/multiple birth cohort. Genet Epidemiol 2014 Dec 22;38(8):692-698 [FREE Full text] [doi: 10.1002/gepi.21855] [Medline: 25250975]
4. Van Schalkwyk GJ, Tran E, Chang K. The impact of Macao’s gaming industry on family life. China Perspectives 2006 Apr 01:1. [doi: 10.4000/chinaperspectives.603]

5. Wang K, Gaitsch H, Poon H, Cox NJ, Rzhetsky A. Classification of common human diseases derived from shared genetic and environmental determinants. Nat Genet 2017 Sep;49(9):1319-1325 [FREE Full text] [doi: 10.1038/ng.3931] [Medline: 28783162]
6. Polubriaginof FCG, Vanguri R, Quinnies K, Belbin GM, Yahi A, Salmasian H, et al. Disease heritability inferred from familial relationships reported in medical records. Cell 2018 Jun 14;173(7):1692-1704 [FREE Full text] [doi: 10.1016/j.cell.2018.04.032] [Medline: 29779949]
7. Tourassi G, Yoon H, Xu S. A novel web informatics approach for automated surveillance of cancer mortality trends. J Biomed Inform 2016 Jun;61:110-118 [FREE Full text] [doi: 10.1016/j.jbi.2016.03.027] [Medline: 27044930]
8. Yoon H, Tourassi G, Xu S. Residential mobility and lung cancer risk: data-driven exploration using internet sources. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 2015;9021:464-469 [FREE Full text] [doi: 10.1007/978-3-319-16268-3_60] [Medline: 26618206]
9. He K, Wu J, Ma X, Zhang C, Huang M, Li C, et al. Extracting kinship from obituary to enhance electronic health records for genetic research. 2019 Presented at: Proceedings of the Fourth Social Media Mining for Health Applications (#SMM4H) Workshop & Shared Task; 2019; Florence p. 1-10 URL: https://aclanthology.org/W19-3201.pdf [doi: 10.18653/v1/w19-3201]
10. Caruana R. Multitask learning. Mach Learn 1997;28(1):41-75. [doi: 10.1007/978-1-4615-5529-2_5]
11. Collobert R, Weston J. A unified architecture for natural language processing. Proc 25th Int Conf Mach Learn 2008:160-167. [doi: 10.1145/1390156.1390177]
12. Li F, Zhang M, Fu G, Ji D. A neural joint model for entity and relation extraction from biomedical text. BMC Bioinformatics 2017 Mar 31;18(1):198 [FREE Full text] [doi: 10.1186/s12859-017-1609-9] [Medline: 28359255]
13. Verga P, Strubell E, McCallum A. Simultaneously self-attending to all mentions for full-abstract biological relation extraction. 2018 Presented at: Proc 2018 Conf North Am Chap Assoc Comput Linguist; 2018; New Orleans p. 872-884. [doi: 10.18653/v1/n18-1080]
14. Sun C, Gong Y, Wu Y, Gong M, Jiang D, Lan M, et al. Joint type inference on entities and relations via graph convolutional networks. 2020 Presented at: Proc 57th Ann Mtg Assoc Comput Linguist; 2020; Florence p. 1361-1370. [doi: 10.18653/v1/p19-1131]
15. Bhatia P, Celikkaya B, Khalilia M. Joint entity extraction and assertion detection for clinical text. 2019 Presented at: Proc 57th Ann Mtg Assoc Comput Linguist; 2019; Florence p. 954-959. [doi: 10.18653/v1/p19-1091]
16. Zheng S, Wang F, Bao H, Hao Y, Zhou P, Xu B. Joint extraction of entities and relations based on a novel tagging scheme. 2017 Presented at: Proc 55th Ann Mtg Assoc Comput Linguist; 2017; Vancouver p. 1227-1236 URL: https://aclanthology.org/P17-1113.pdf [doi: 10.18653/v1/p17-1113]
17. Dixit K, Al-Onaizan Y. Span-level model for relation extraction. 2019 Presented at: Proc 57th Ann Mtg Assoc Comput Linguist; 2019; Florence p. 5308-5314 URL: https://aclanthology.org/P19-1525.pdf [doi: 10.18653/v1/p19-1525]
18. Zheng S, Hao Y, Lu D, Bao H, Xu J, Hao H, et al. Joint entity and relation extraction based on a hybrid neural network. Neurocomputing 2017 Sep;257:59-66. [doi: 10.1016/j.neucom.2016.12.075]
19. Wei J, Zou K. EDA: easy data augmentation techniques for boosting performance on text classification tasks. 2019 Conf Empir Methods Nat Lang Process 9th Int Jt Conf Nat Lang Process Proc Conf Association for Computational Linguistics 2020:6382-6388 [FREE Full text] [doi: 10.18653/v1/d19-1670]
20. St Sauver JL, Grossardt BR, Yawn BP, Melton LJ, Rocca WA. Use of a medical records linkage system to enumerate a dynamic population over time: the Rochester epidemiology project. Am J Epidemiol 2011 May 01;173(9):1059-1068 [FREE Full text] [doi: 10.1093/aje/kwq482] [Medline: 21430193]
21. Clark K, Luong M, Manning C, Le QV. Semi-supervised sequence modeling with cross-view training. 2018 Presented at: Proc 2018 Conf Empir Methods Nat Lang Proc; 2018; Brussels p. 1914-1925. [doi: 10.18653/v1/d18-1217]
22. Devlin J, Chang M, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding (bidirectional encoder representations from transformers). 2019 Presented at: NAACL-HLT; 2019; North Stroudsburg p. 4171-4186 URL: https://nlp.stanford.edu/seminar/details/jdevlin.pdf [doi: 10.18653/v1/N19-1423]
23. Lee K, He L, Lewis M, Zettlemoyer L. End-to-end neural coreference resolution. 2017 Presented at: EMNLP 2017 Proc 2017 Conf Empir Methods Nat Lang Proc; 2017; Copenhagen p. 188-197 URL: https://aclanthology.org/D17-1018.pdf [doi: 10.18653/v1/d17-1018]
24. Liu S, Liang Y, Gitter A. Loss-balanced task weighting to reduce negative transfer in multi-task learning. Proc AAAI Conf Artific Intell 2019 Jul 17;33(01):9977-9978 [FREE Full text] [doi: 10.1609/aaai.v33i01.33019977]
25. Wang Z, Dai Z, Poczos B, Carbonell J. Characterizing and avoiding negative transfer. 2019 Presented at: Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit IEEE Computer Society; .?11294; 2019; Long Beach p. 11285-11294 URL: https://openaccess.thecvf.com/content_CVPR_2019/papers/Wang_Characterizing_and_Avoiding_Negative_Transfer_CVPR_2019_paper.pdf [doi: 10.1109/cvpr.2019.01155]
26. Bender J, Cyr AB, Arbuckle L, Ferris LE. Ethics and privacy implications of using the internet and social media to recruit participants for health research: a privacy-by-design framework for online recruitment. J Med Internet Res 2017 Apr 06;19(4):e104 [FREE Full text] [doi: 10.2196/jmir.7029] [Medline: 28385682]
27. Denecke K, Bamidis P, Bond C, Gabarron E, Househ M, Lau AYS, et al. Ethical issues of social media usage in healthcare. Yearb Med Inform 2015 Aug 13;10(1):137-147 [FREE Full text] [doi: 10.15265/IY-2015-001] [Medline: 26293861]

28. Jamal M, Qi G. Task agnostic meta-learning for few-shot learning. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit. 2019. URL: https://openaccess.thecvf.com/content_CVPR_2019/papers/Jamal_Task_Agnostic_Meta-Learning_for_Few-Shot_Learning_CVPR_2019_paper.pdf [accessed 2021-07-13]
29. Gao T, Han X, Zhu H, Liu Z, Li P, Sun M, et al. FewRel 2.0: towards more challenging few-shot relation classification. 2019 Presented at: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP); 2019; Hong Kong p. 6251-6256. [doi: 10.18653/v1/d19-1649]
30. Lin Y, Chen C, Lee J, Li Z, Zhang Y, Xia M, et al. Choosing transfer languages for cross-lingual learning. 2019 Presented at: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics; 2019; Florence p. 3125-3135 URL: https://aclanthology.org/P19-1301.pdf [doi: 10.18653/v1/p19-1301]
31. Ruder S, Søgaard A, Vuli I. Unsupervised cross-lingual representation learning. 2019 Presented at: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts; 2019; Florence p. 31-38 URL: https://aclanthology.org/P19-4007.pdf [doi: 10.18653/v1/p19-4007]
32. Pan SJ, Yang Q. A survey on transfer learning. IEEE Trans Knowl Data Eng 2010 Oct;22(10):1345-1359 [FREE Full text] [doi: 10.1109/tkde.2009.191]
33. Genealogical knowledge graph. URL: https://github.com/KaiHe-better/Genealogical-Knowledge-Graph [accessed 2021-07-19]

Abbreviations

EHR: electronic health records
GKG: genealogical knowledge graph
LSTM: long short-term memory
NER: name entity recognition
NLP: natural language processing
PII: personally identifiable information
RC: relation classification

Edited by R Kukafka; submitted 11.11.20; peer-reviewed by Z Ben-Miled, H Salmasian, M Torii; comments to author 15.12.20; revised version received 26.01.21; accepted 21.06.21; published 04.08.21.

Please cite as:
He K, Yao L, Zhang J, Li Y, Li C
Construction of Genealogical Knowledge Graphs From Obituaries: Multitask Neural Network Extraction System
J Med Internet Res 2021;23(8):e25670
URL: https://www.jmir.org/2021/8/e25670
doi:10.2196/25670
PMID:34346903

©Kai He, Lixia Yao, JiaWei Zhang, Yufei Li, Chen Li. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 04.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

An Artificial Neural Network Prediction Model for Posttraumatic Epilepsy: Retrospective Cohort Study

Xueping Wang[1], MD; Jie Zhong[2], MD; Ting Lei[3], MD; Deng Chen[1], MD; Haijiao Wang[1], MD; Lina Zhu[1], MD; Shanshan Chu[1], PhD; Ling Liu[1], MD

[1]Department of Neurology, West China Hospital, Sichuan University, Chengdu, China
[2]Department of Ophthalmology, Sichuan Provincial People’s Hospital, Chengdu, China
[3]Department of Neurosurgery, Shang Jin Nan Fu Hospital of West China Hospital, Sichuan University, Chengdu, China

Corresponding Author:
Ling Liu, MD
Department of Neurology
West China Hospital
Sichuan University
No. 37, Guo Xue Xiang
Chengdu, 610041
China
Phone: 86 151 1705 8487
Email: zjllxx1968@163.com

Abstract

Background: Posttraumatic epilepsy (PTE) is a common sequela after traumatic brain injury (TBI), and identifying high-risk patients with PTE is necessary for their better treatment. Although artificial neural network (ANN) prediction models have been reported and are superior to traditional models, the ANN prediction model for PTE is lacking.

Objective: We aim to train and validate an ANN model to anticipate the risks of PTE.

Methods: The training cohort was TBI patients registered at West China Hospital. We used a 5-fold cross-validation approach to train and test the ANN model to avoid overfitting; 21 independent variables were used as input neurons in the ANN models, using a back-propagation algorithm to minimize the loss function. Finally, we obtained sensitivity, specificity, and accuracy of each ANN model from the 5 rounds of cross-validation and compared the accuracy with a nomogram prediction model built in our previous work based on the same population. In addition, we evaluated the performance of the model using patients registered at Chengdu Shang Jin Nan Fu Hospital (testing cohort 1) and Sichuan Provincial People’s Hospital (testing cohort 2) between January 1, 2013, and March 1, 2015.

Results: For the training cohort, we enrolled 1301 TBI patients from January 1, 2011, to December 31, 2017. The prevalence of PTE was 12.8% (166/1301, 95% CI 10.9%-14.6%). Of the TBI patients registered in testing cohort 1, PTE prevalence was 10.5% (44/421, 95% CI 7.5%-13.4%). Of the TBI patients registered in testing cohort 2, PTE prevalence was 6.1% (25/413, 95% CI 3.7%-8.4%). The results of the ANN model show that, the area under the receiver operating characteristic curve in the training cohort was 0.907 (95% CI 0.889-0.924), testing cohort 1 was 0.867 (95% CI 0.842-0.893), and testing cohort 2 was 0.859 (95% CI 0.826-0.890). Second, the average accuracy of the training cohort was 0.557 (95% CI 0.510-0.620), with 0.470 (95% CI 0.414-0.526) in testing cohort 1 and 0.344 (95% CI 0.287-0.401) in testing cohort 2. In addition, sensitivity, specificity, positive predictive values and negative predictors in the training cohort (testing cohort 1 and testing cohort 2) were 0.80 (0.83 and 0.80), 0.86 (0.80 and 0.84), 91% (85% and 78%), and 86% (80% and 83%), respectively. When calibrating this ANN model, Brier scored 0.121 in testing cohort 1 and 0.127 in testing cohort 2. Compared with the nomogram model, the ANN prediction model had a higher accuracy (P=.01).

Conclusions: This study shows that the ANN model can predict the risk of PTE and is superior to the risk estimated based on traditional statistical methods. However, the calibration of the model is a bit poor, and we need to calibrate it on a large sample size set and further improve the model.

(J Med Internet Res 2021;23(8):e25090) doi:10.2196/25090

KEYWORDS

artificial neural network; posttraumatic epilepsy; traumatic brain injury

Introduction

Background

Traumatic brain injuries (TBIs) reduce patient quality of life and result in high morbidity and mortality [1]. TBI can also lead to a range of sequelae, the most common being posttraumatic epilepsy (PTE), which accounts for 4% to 9% of all epilepsy cases [2-5]. Population-based and cohort studies estimate the overall incidence of PTE ranges from 5% to 50%, especially among war veterans, who receive more penetrating TBIs than civilians [6-12]. Previous literature concludes that the incidence of PTE increases with the severity of TBI [5-7], and the vast majority of PTE appears within the first 2 years after TBI and rises in the following 30 years [3,13].

Because of the high incidence and adverse effects of PTE, clinicians need to identify and better manage those patients at high risk. PTE risk factors such as TBI severity, brain contusion, subdural hematoma, neurosurgery, and early posttraumatic seizure (PTS) are reported by multiple regression methods, etc [14-17]. These results are expressed as risk ratios or odds ratios, but they are inconvenient to use. A lot of TBI patients received antiepileptic prophylaxis to prevent PTE. While clinical trials have shown that antiepileptic prophylaxis within 7 days of TBI reduces the incidence of early seizure attacks, a reduction in PTE has not been seen [18,19]. The negative results of these studies may be due to blind selection of the study population and insufficient follow-up time. In addition, antiepileptic prophylaxis for patients with low risk would add financial burden and side effects, so it is necessary for clinicians to identify those at high risk of PTE. However, so far there are no reliable tools to predict the risk of PTE; if we can predict this via the web, it will be of great significance in the realization of precision medicine [20]. Artificial neural network (ANN) is a form of artificial intelligence that can mimic the problem-solving process of the human brain and generate a mathematical algorithmic model that can handle the nonlinear relationship between variables [21]. ANN is one of the most commonly used methods of supervising machine learning, consisting of 3 layers of neurons: an input layer of independent variables, a hidden layer for processing information, and an output layer for the probability of an outcome. ANN-based risk predictive models have several advantages; they can capture nonlinear relationships among input variables, making them ideal candidates for classifying complex diseases [22,23] such as identifying high-risk transient ischemic attack or minor stroke [24] and assisting in precision medicine for COVID-19 [25]. Compared with logistic regression models, ANN models can predict a complex relationship between variables and are more accurate in classifying the dependent variable [26].

Aim and Research Questions

To date, no published papers have focused on predicting PTE after TBI through ANN. To investigate this problem, we applied ANN to analyze demographic, clinical, and radiological data from TBI patients to achieve accurate prediction of PTE for individual patients, thereby recognizing PTE patients as early as possible, which might be helpful for further antiepileptogenic intervention studies through identifying the suitable target population. In our previous study, we had built a nomogram model to predict PTE based on the same population, and we wondered whether the ANN model outperformed it.

Methods

Design

The study had a retrospective cohort design, and the West China Hospital of Sichuan University Ethics Committee approved this study (no. 2019-936). Subjects or their proxies gave informed verbal consent to participate in this study.

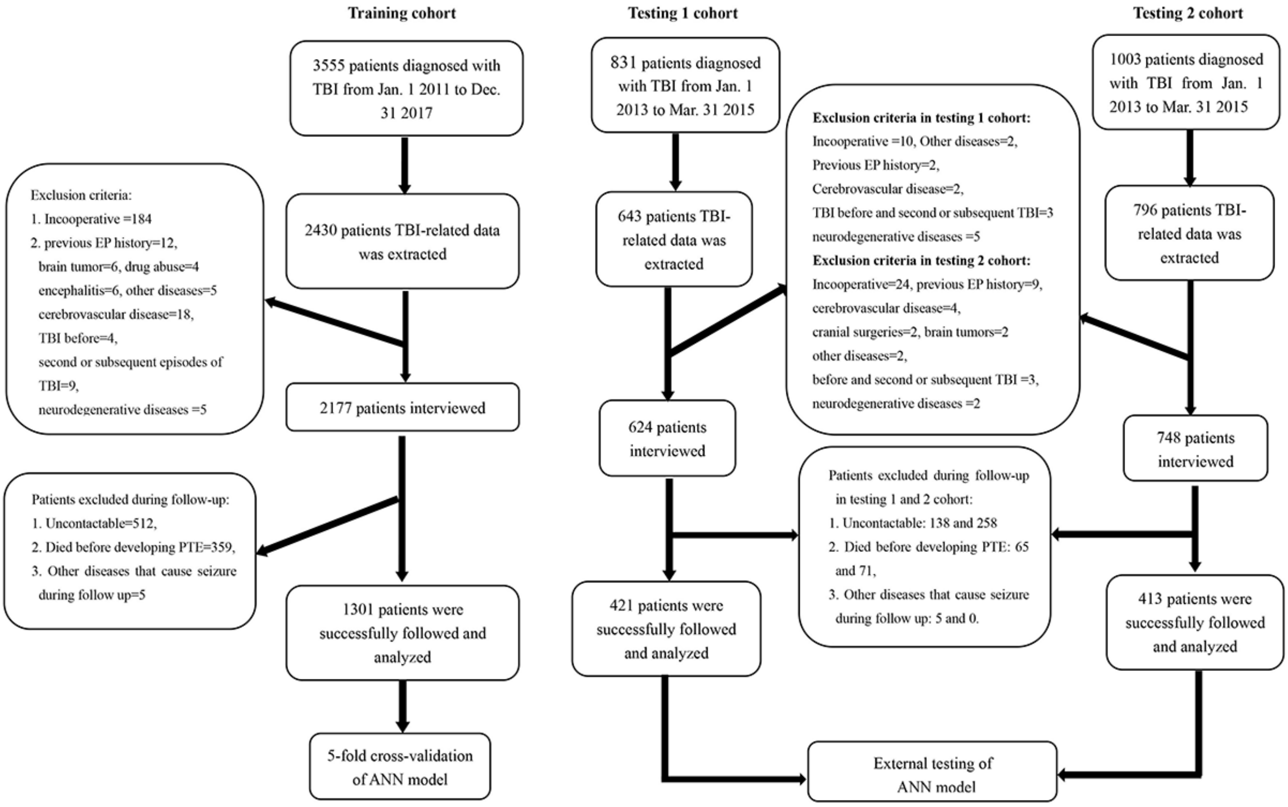
Participants

This ANN predictive model was developed on a retrospective registry of TBI patients at the West China Hospital (a tertiary referral center in Sichuan province, China) from January 1, 2011, to December 31, 2017. These subjects were the training cohort. The model was also tested in 2 external cohorts registered at Chengdu Shang Jin Nan Fu Hospital (testing cohort 1, n=421) and Sichuan Provincial People's Hospital (testing cohort 2, n=413) between January 1, 2013, and March 1, 2015. All patients were diagnosed with TBI, which was defined as any hospital discharge with a primary or secondary diagnosis of trauma to the head. According to the *International Statistical Classification of Diseases and Related Health Problems, Tenth Revision* (ICD-10), patient records of those diagnosed with traumatic brain injury (S06.902), cerebral concussion (S06.001), subdural hematoma (S06.501), epidural hematoma (S06.401), traumatic subarachnoid hemorrhage (S06.601), skull fracture (liner or depressed fracture; S02.902), traumatic intracranial hemorrhage (S06.806), brain contusion (S06.201), diffuse axonal injury (S06.204), and open or closed TBI (S06.911) were extracted from the electronic medical record database.

We included all patients with complete demographic, clinical, and radiological data to determine TBI, PTS, and PTE.

The inclusion criteria were (1) brain injury was caused by an external force, (2) clinical diagnosis of TBI, (3) TBI occurred between January 1, 2011, and December 31, 2017, for West China Hospital patients and for Chengdu Shang Jin Nan Fu Hospital and Sichuan Provincial People's Hospital patients from January 1, 2013, to March 1, 2015, (4) complete trauma-related data were available in medical records, and (5) patients or their relatives agreed to participate in this study (Figure 1).

Figure 1. Study cohort. ANN: artificial neural network; TBI: traumatic brain injury


Training cohort
Testing 1 cohort
Testing 2 cohort
3555 patients diagnosed with TBI from Jan. 1 2011 to Dec. 31 2017
831 patients diagnosed with TBI from Jan. 1 2013 to Mar. 31 2015
1003 patients diagnosed with TBI from Jan. 1 2013 to Mar. 31 2015
Exclusion criteria: 1. Incooperative =184 2. previous EP history=12, brain tumor=6, drug abuse=4 encephalitis=6, other diseases=5 cerebrovascular disease=18, TBI before=4, second or subsequent episodes of TBI=9, neurodegenerative diseases =5
2430 patients TBI-related data was extracted
643 patients TBI-related data was extracted
Exclusion criteria in testing 1 cohort: Incooperative =10, Other diseases=2, Previous EP history=2, Cerebrovascular disease=2, TBI before and second or subsequent TBI=3 neurodegenerative diseases =5
Exclusion criteria in testing 2 cohort: Incooperative=24, previous EP history=9, cerebrovascular disease=4, cranial surgeries=2, brain tumors=2 other diseases=2, before and second or subsequent TBI =3, neurodegenerative diseases =2
796 patients TBI-related data was extracted
2177 patients interviewed
624 patients interviewed
748 patients interviewed
Patients excluded during follow-up: 1. Uncontactable=512, 2. Died before developing PTE=359, 3. Other diseases that cause seizure during follow up=5
Patients excluded during follow-up in testing 1 and 2 cohort: 1. Uncontactable: 138 and 258 2. Died before developing PTE: 65 and 71, 3. Other diseases that cause seizure during follow up: 5 and 0.
1301 patients were successfully followed and analyzed
421 patients were successfully followed and analyzed
413 patients were successfully followed and analyzed
5-fold cross-validation of ANN model
External testing of ANN model

The following patients were excluded: (1) patients who had epilepsy or seizures before TBI; (2) patients who had a previous TBI or had second or subsequent episodes of TBI; (3) patients with other conditions that can cause seizure, such as cerebrovascular disease, brain tumors, encephalitis, brain surgery, and other chronic diseases; (4) patients whose general condition was poor or who had other conditions that may lead to epileptic seizures before PTS or PTE came out during follow-up.

Data and Data Collection

Formally trained neurologists extracted the necessary data for model building from the hospital records of patients. The table included the general condition (age, sex, length of hospital stay, previous history), the clinical and radiological data of TBI (mechanism of TBI, severity of TBI, clinical manifestations, treatments, brain CT performed at initial presentation), and the seizure onset information during their hospitalization (early PTS and immediate PTS). Variables used to construct the predictive model and how we defined and classified them are listed in Table 1.

With the new definition proposed by the International League Against Epilepsy and the International Bureau for Epilepsy, epilepsy requires at least 2 unprovoked (or reflex) seizures occurring more than 24 hours apart and one unprovoked (or reflex) seizure and a probability of further seizures similar to the general recurrence risk (at least 60%) after 2 unprovoked seizures, occurring over the next 10 years [27]. PTS was defined as a single, nonrecurrent convulsive episode that fits in 1 of 3 categories according to the time of seizure onset: immediate PTS, occurring in the first 24 hours following injury; early PTS, occurring more than 24 hours following injury and within 7 days; and late PTS, occurring more than 1 week after trauma. In our study, PTE refers to one or more recurrent seizures occurring more than 1 week after TBI, including late PTS.

Table 1. Comparison of demographic data between the posttraumatic epilepsy and non–posttraumatic epilepsy patients in the training cohort.

Variable	Total (n=1301)	Non-PTE[a] (n=1135)	PTE (n=166)	*P* value
Demographic data	—[b]	—	—	—
Sex, male, n (%)	983 (75.6)	850 (74.6)	133 (82.6)	.03
Age (years), mean (SD)	38.12 (23.82)	37.79 (24.4)	40.41 (19.09)	.19
≤15, n (%)	297 (22.8)	77 (24.3)	20 (12.4)	<.01
16-40, n (%)	358 (27.5)	304 (26.7)	54 (33.5)	—
41-64, n (%)	447 (34.4)	373 (32.7)	74 (46.0)	—
≥65, n (%)	199 (15.3)	186 (16.3)	13 (8.1)	—
Follow-up (months), mean (SD)	71.49 (22.54)	72.37 (21.98)	65.6 (25.5)	.08
Clinical characteristics	—	—	—	—
GCS[c] score, mean (SD)	11.72 (3.8)	12.22 (3.58)	8.17 (3.40)	.008
13-15, n (%)	794 (61)	768 (67.4)	26 (16.1)	<.001
9-12, n (%)	227 (17.4)	190 (16.7)	37 (23)	—
3-8, n (%)	280 (21.5)	182 (16)	98 (60.9)	—
LOH[d] (days), mean (SD)	14 (23.64)	12.31 (22.17)	25.91 (29.63)	<.001
Neurological deficits, n (%)	386 (29.7)	293 (25.7)	93 (57.8)	<.001
LOC[e], n (%)	666 (51.2)	545 (47.8)	121 (75.2)	<.001
LOC time, n (%)	—	—	—	<.001
0-30 minutes	812 (62.4)	782 (68.6)	30 (18.6)	—
31 minutes-24 hours	113 (8.7)	96 (8.4)	17 (10.6)	—
>24 hours	376 (28.9)	262 (23)	114 (70.8)	—
Etiology of TBI[f] , n (%)	—	—	—	.06
MVA[g]	529 (40.7)	445 (39)	84 (52.5)	—
Violence	135 (10.4)	121 (10.6)	14 (8.7)	—
Fall ≤1 m	376 (28.9)	341 (29.9)	35 (21.7)	—
Fall >1 m	261 (20.1)	233 (20.4)	28 (17.4)	—
Treatment, n (%)	—	—	—	<.001
Conservative	651 (50)	602 (52.8)	49 (30.4)	—
Neurological surgery	572 (44)	468 (41.1)	104 (64.6)	—
Puncture	78 (6)	78 (6)	8 (5)	—
Neuroimaging results, n (%)	—	—	—	—
SDH[h]	548 (42.1)	455 (39.9)	93 (57.8)	<.001
EDH[i]	422 (32.4)	386 (33.9)	36 (22.4)	.004
ICH[j]	410 (31.5)	318 (27.9)	92 (57.1)	<.001
SAH[k]	336 (25.8)	280 (24.6)	56 (34.8)	.04
DAI[l]	95 (7.3)	68 (6)	27 (16.8)	<.001
Contusion site, n (%)	—	—	—	<.001
None	720 (55.3)	679 (59.6)	41 (25.5)	—
Frontotemporal lobe	240 (18.4)	160 (14)	80 (49.7)	—
Other	341 (26.2)	301 (26.4)	40 (24.8)	—
Fracture, n (%)	—	—	—	.005

Variable	Total (n=1301)	Non-PTE[a] (n=1135)	PTE (n=166)	*P* value
No	704 (54)	634 (55.6)	70 (43.5)	—
Liner	487 (37.4)	408 (35.8)	79 (49.1)	—
Depressed	110 (8.5)	98 (8.6)	12 (7.5)	—
Open	79 (6.1)	68 (6)	11 (6.8)	.60
PTS[m] and PTE, n (%)	—	—	—	—
IPTS[n]	13 (1)	11 (1)	2 (1.2)	.87
EPTS[o]	74 (5.7)	39 (3.4)	35 (21.7)	<.001
PTE	166 (12.8)	—	—	—
Risk of PTE within 1 year	97 (7.5)	—	—	—
Risk of PTE within 5 years	145 (11.1)	—	—	—
Risk of PTE within 8 years	166 (12.8)	—	—	—
PTE in mild TBI	28 (2.2)	—	—	—
PTE in moderate TBI	39 (3.0)	—	—	—
PTE in severe TBI	99 (7.6)	—	—	—

[a] PTE: posttraumatic epilepsy.
[b] Not applicable.
[c] GCS: Glasgow Coma Scale.
[d] LOH: length of hospital stay.
[e] LOC: loss of consciousness.
[f] TBI: traumatic brain injury.
[g] MVA: motor vehicle accident.
[h] SDH: subdural hematoma.
[i] EDH: epidural hematoma.
[j] ICH: intracranial hemorrhage.
[k] SAH: subarachnoid hemorrhage.
[l] DAI: diffuse axonal injury.
[m] PTS: posttraumatic epilepsy.
[n] IPTS: immediate posttraumatic seizure.
[o] EPTS: early posttraumatic seizure.

Follow-Up and Data Collection

All participants were followed for at least 1 year to monitor for seizures. Among the participants who were unable to understand the survey, we interviewed their close relatives and their general practitioner. The follow-up investigations contain the general condition of the patient (whether there was cachexia or other diseases that can lead to misdiagnosis of PTE), the occurrence of seizures (when did the first seizure attack appear after discharge from the hospital), the type and frequency of seizures (the clinical manifestations and frequency of epileptic seizures), and the treatment condition (whether they took antiepileptic drugs and the drug dosage). If patients or their caregivers reported a seizure attack, neurologists in our team would interview them face-to-face and determined the diagnosis by the definition of PTE according to their clinical manifestations and electroencephalogram results. The main outcome measure was the incidence of PTE.

Statistical Analysis

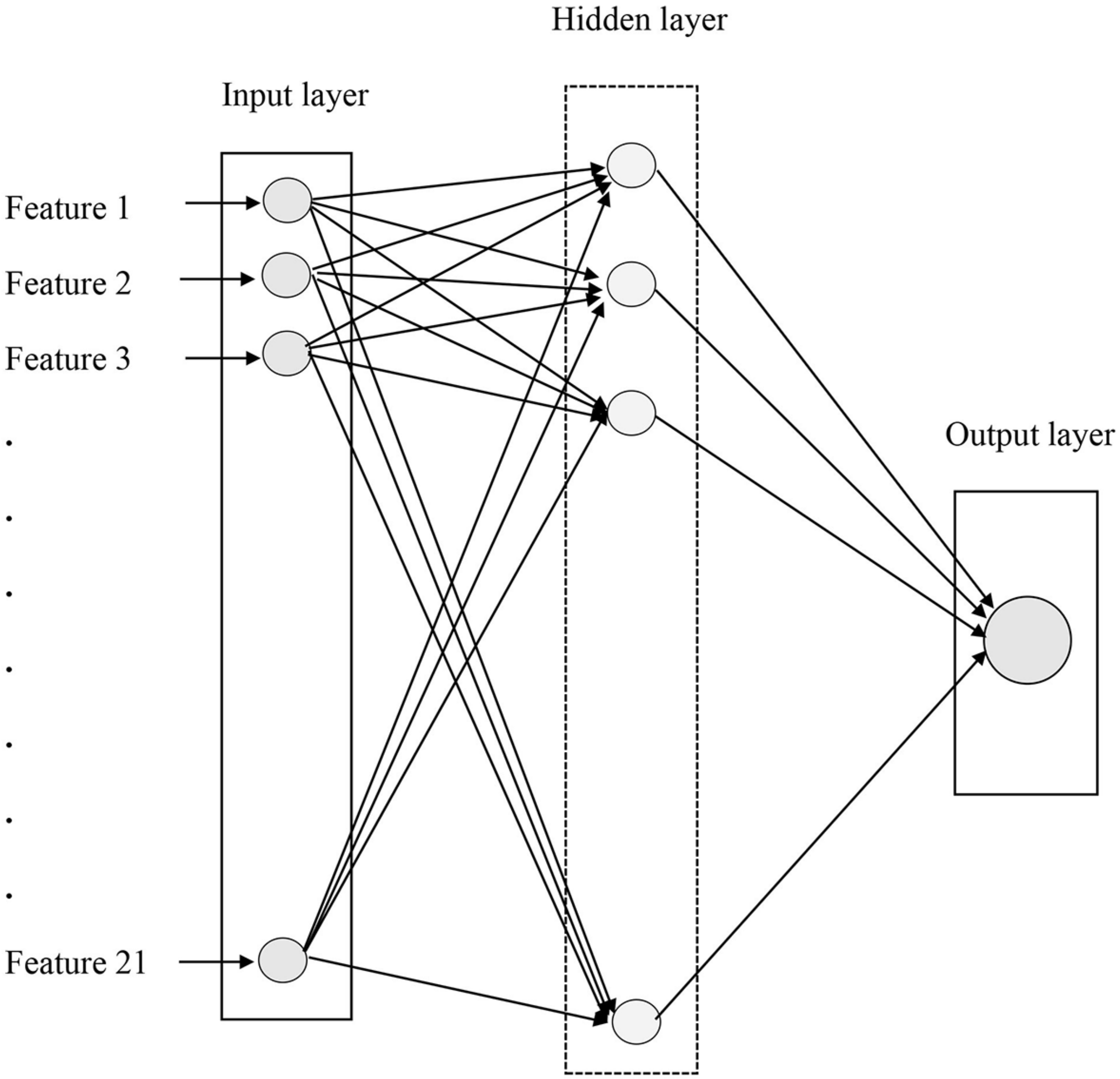
ANN Model

The most common 3-layer multilayer perceptron ANN model was employed in this study (Figure 2). The input layer incorporated 21 independent variables (Table 1). We performed 5 rounds of model learning and validation (step 1) and calculated the average area under the curve (AUC) using the results of 5 model validations, which represented an estimate of the accuracy of the model. Model test (step 2) was performed in another 2 sets.

Step 1: Model Development and Validation

We developed the ANN model using the Keras framework with Python 3.6 programming language (Python Software Foundation). The learning algorithm was back-propagation. Back-propagation can minimize the loss function by iteratively updating the weights between neurons, maximizing the predictive power of the ANN model for the main results. We constructed a cost-sensitive support vector machine classification prediction mode by setting those factors related to PTE as input variables and PTE as an output variable. Given the small sample size, we used a 5-fold cross-validation to validate the ANN models to avoid overfitting [28,29]. The training dataset was randomly divided into 5 folds, and we performed 5 rounds of training and validation of the ANN models. During the 5-fold cross-validation process, 4 folds were the training subsets and the remaining fold was the validation subset, each of which was used only once as a validation set. We first obtained the model index from 5-fold cross-validation, selected the hyperparameters through training, and then retrained the full amount of training data with the optimal parameters of the optimal model. After many attempts, we finally identified 43 (2 × 21 + 1) hidden neurons with 5000 training rounds to train the entire training set. The model began to enter the overfitting phase when the number of training rounds exceeded 5000 epochs.

Figure 2. Optimal network architecture of the artificial neural network: a multilayer perceptron.



Step 2: Model Test

The predictive performance of the final ANN model was evaluated using 2 external testing datasets (Shang Jin Nan Fu Hospital and Sichuan Provincial People's Hospital) unknown to the training models. The testing datasets were used for final model evaluation after cross-validation process, model fit, and probability calibration. After setting the ANN hyperparameters, we started to train the neural network with the full amount of training data and stopped training after reaching 5000 times. The test datasets are predicted for about 50 times using a trained neural network, and the results of each prediction were recorded.

Evaluating Prediction Accuracy

The performance of the ANN model was measured by its accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). Since the primary outcome was a binary variable (PTE or not), area under the receiver operating characteristic curve (ROC), referred to as AUC, was used to assess the accuracy of this predictive model.

The average precision, equivalent to the area under the precision-recall curve, was measured for evaluating model performance. The 95% confidence intervals were determined by 50 rounds of model testing (random sampling in the test dataset, learning and verification, and then repeat multiple times). Model calibration was assessed in 2 testing cohorts by calculating Brier scores to examine how well the model predicting PTE frequencies matched the observed one.

Statistical analysis was performed using SPSS (version 22.0, IBM Corp). Independent sample *t* tests were used to compare quantitative data with a normal distribution; otherwise the Mann-Whitney *U* test was applied. The results were presented as mean and standard deviation or interquartile range. The incidence rates were expressed in percentile; to examine associations of categorical and quantitative prognostic factors with the development of PTE, the Fisher exact test and Mann-Whitney *U* test were applied, respectively. In our previous work, by using the rms package in R (version 3.5.1, R Foundation for Statistical Computing), a nomogram was formulated with 7 independent risk factors of PTE founded with multivariate Cox proportional hazards regression analysis based on the cohort of West China Hospital. We compared the prediction accuracy of ANN model with nomogram model [30] using a DeLong test. *P* values reported are 2-tailed, and a value *P*<.05 was considered significant.

Results

Patient Characteristics

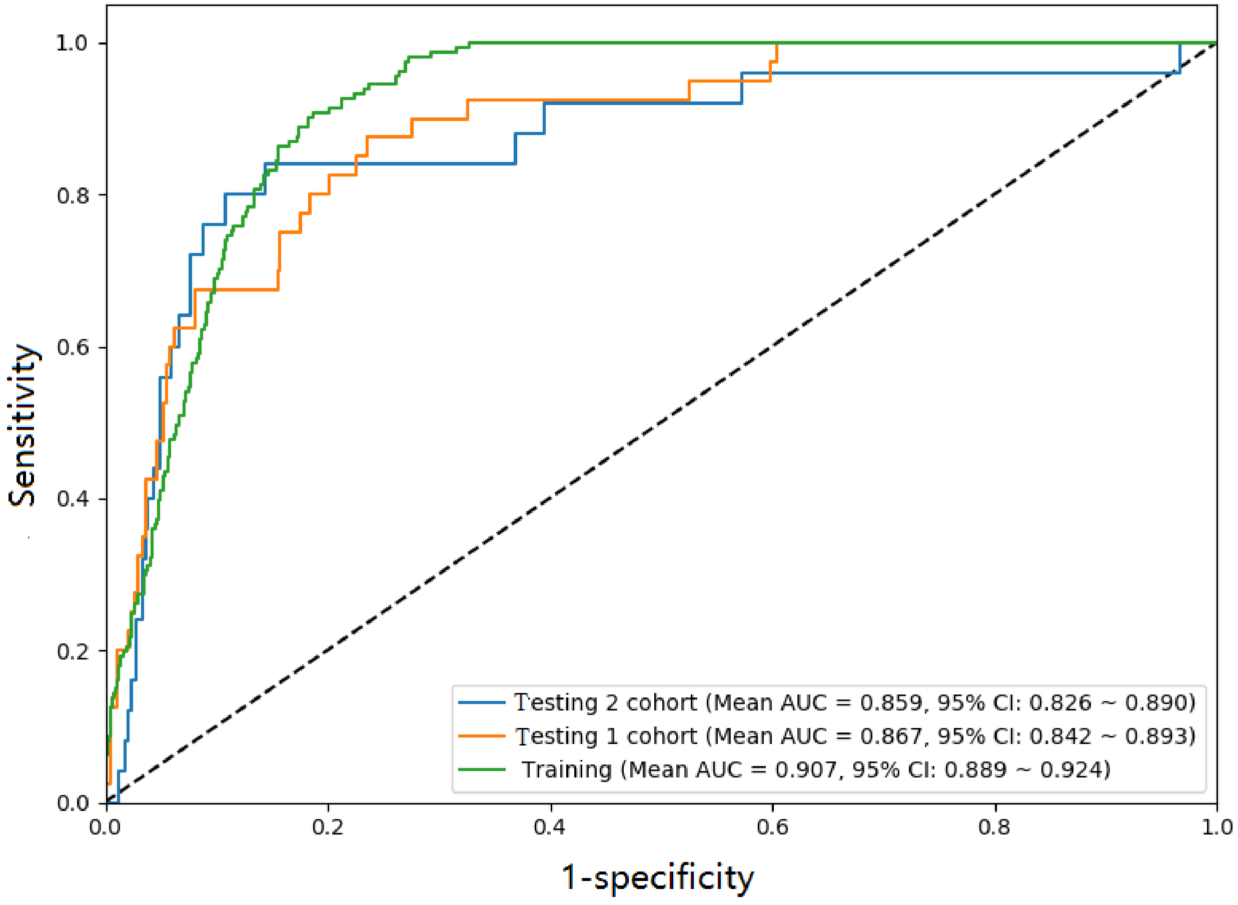
A total of 2135 patients were included in this study. Between January 1, 2011, and December 31, 2017, 1301 subjects from West China Hospital were enrolled as the training cohort, and the prevalence of PTE was 12.8% (166/1301, 95% CI 10.9%-14.6%). From January 1, 2013, to March 1, 2015, 421 patients from Shang Jin Nan fu Hospital were testing cohort 1, and the prevalence of PTE was 10.5% (44/421, 95% CI 7.5%-13.4%). A total of 413 patients from Sichuan Provincial People’s Hospital were testing cohort 2, and the prevalence of PTE was 6.1% (25/413, 95% CI 3.7%-8.4%). The prevalence of PTE among 3 cohorts had significant difference (*P*=.001).

The comparison of demographic data, clinical manifestation, and radiological results between the PTE- and non-PTE groups in the training cohort was listed in Table 1. Significant differences were found in many variables, including sex, age group, length of hospital days, etiology of TBI, loss of consciousness time, treatment, subdural hematoma, intracranial hemorrhage, diffuse axonal injury, contusion load, contusion site, fracture, and early PTS (both *P*<.001 for all variables). There was no significant difference in follow-up time, subarachnoid hemorrhage, epidural hematoma, open TBI, and intermediate PTS between patients with PTE and non-PTE.

ANN Predictive Model Performance

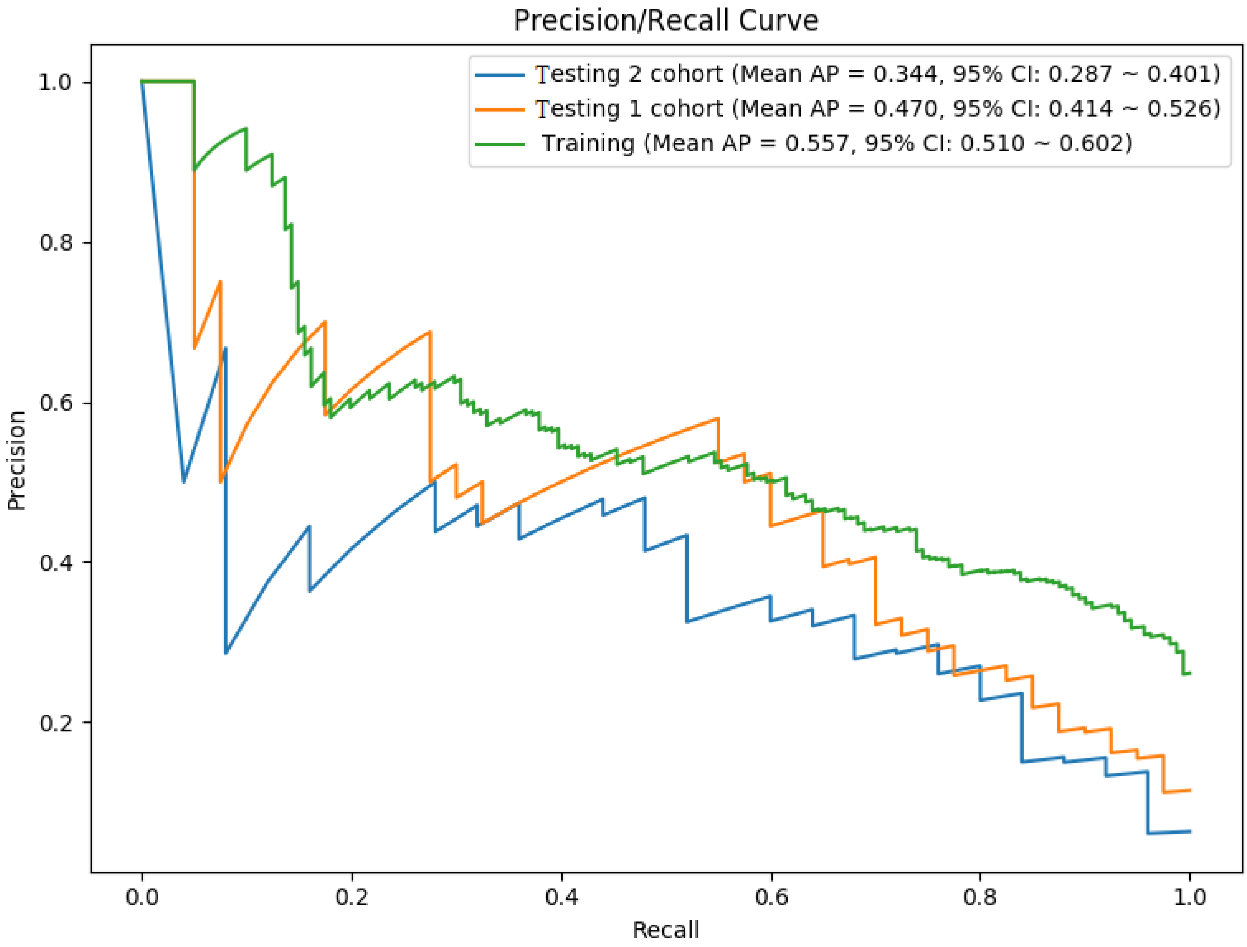
The ANN prediction model incorporated 21 features from each patient in the training cohort to predict whether an individual would develop PTE. In the training cohort, the mean AUC of the ANN model was 0.907 (95% CI 0.889-0.924), the sensitivity and specificity were 0.80 and 0.86, the PPV and NPV were 91% and 86%, which means 91% of patients who developed PTE and 86% of patients who did not develop PTE were exactly predicted by the ANN model. When testing the ANN model with datasets from Shang Jin Nan Fu Hospital and Sichuan Provincial People’s Hospital, the AUCs were 0.867 (95% CI 0.842-0.893) and 0.859 (95% CI 0.826-0.890), sensitivity was 0.83 and 0.80, specificity was 0.80 and 0.84, PPV was 85% and 78%, and NPV was 80% and 83%. The greater the value of the AUC, the better the performance of the model, which means higher predictive accuracy in this study. Figure 3 shows the ROC of the training cohort and 2 testing cohorts.

Figure 3. Evaluation of artificial neural network prediction model accuracy using receiver operating characteristic. ANN: artificial neural network; AUC: area under curve.



In addition, the average precision of the ANN prediction model in training cohort was 0.557 (95% CI 0.510-0.620), while the average precision was 0.470 (95% CI 0.414-0.526) in testing cohort 1 and 0.344 (95% CI 0.287-0.401) in testing cohort 2 (Figure 4). Similar to AUC, the larger the average precision value, the better the prediction accuracy. However, unlike the ROC, the area under the precision-recall curve was less than 0.5, which did not mean that the prediction performance of the model was poor. An asymmetric data distribution (ie, the number of negative, or non-PTE, events is much more than the number of positive, or PTE, events) leads to a low overall decrease in average precision and will have a great impact on the precision-recall curve but no effect on the AUC curve [31].

Figure 4. Evaluation of the artificial neural network prediction model accuracy using precision-recall curves. ANN: artificial neural network; AP: average precision.



Brier scores were calculated to evaluate the calibration of the ANN prediction model. With the Brier score, we can know the calibration of the prediction model [32]. It ranges from 0 to 1; the lower the Brier score, the better the calibration, so the ideal Brier score is 0, indicating the prediction is completely accurate [33]. When testing our ANN prediction model, the Brier scores were 0.121 in testing cohort 1 and 0.127 in testing cohort 2. Figure 5 shows the calibration plots that compare the proportion of PTE patients predicted by the ANN model with the actual observed rate of PTE. The diagonal curve represents a perfectly calibrated prediction, and the calibration curve should be as close to this diagonal curve as possible. In our study, the calibration curves in the 2 testing groups were a little far away from the diagonal curve, and we needed to calibrate this ANN prediction model with a large-sample dataset. Table 2 showed the performances of the ANN model on the training and 2 testing sets.

Figure 5. Risk calibration curves for artificial neural network prediction model in two testing cohorts. A curve closer to the dotted diagonal line indicates better calibration, with corresponding lower Brier score. ANN: artificial neural network.

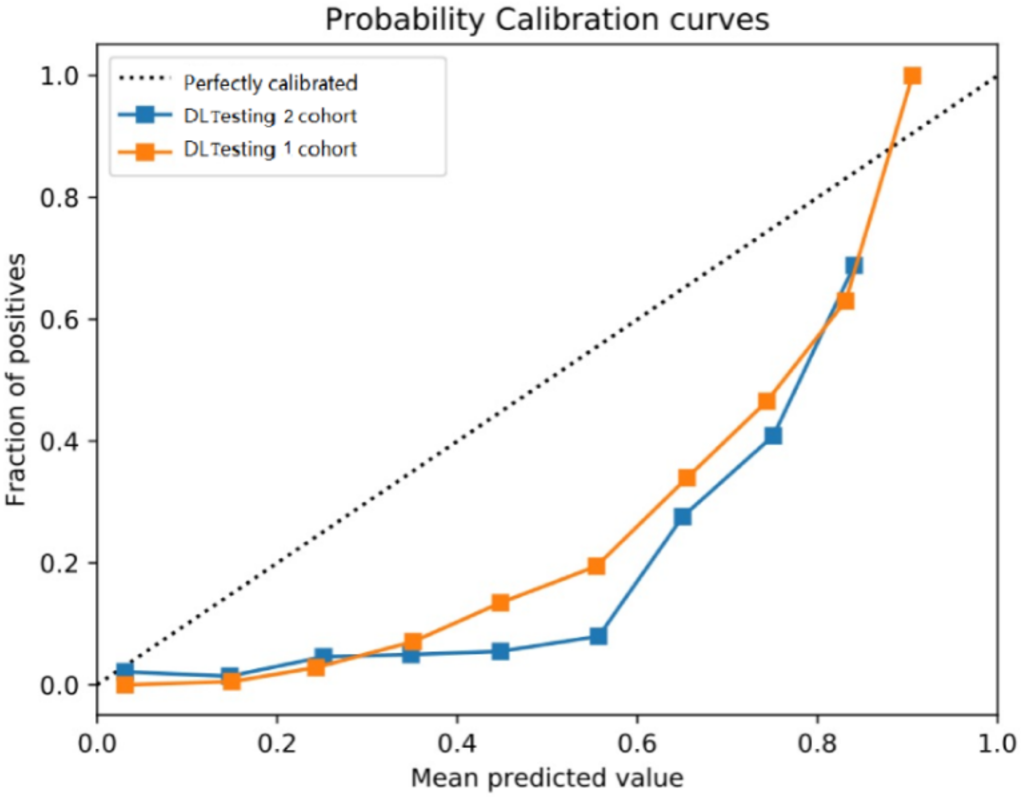


Table 2. Performances of artificial neural network on training and two testing sets.

Indicator	Training cohort	Testing cohort 1	Testing cohort 2
AUC[a] (95% CI)	0.907 (0.889-0.924)	0.867 (0.842-0.893)	0.859 (0.826-0.890)
Sensitivity	0.80	0.83	0.80
Specificity	0.86	0.80	0.84
AP[b] (95% CI)	0.557 (0.510-0.620)	0.470 (0.414-0.526)	0.344 (0.287-0.401)
PPV[c], %	91	85	78
NPV[d], %	86	80	83
Brier score	—[e]	0.121	0.127

[a] AUC: area under the curve.

[b] AP: average precision.

[c] PPV: positive predictive value.

[d] NPV: negative predictive value.

[e] Not applicable.

Compared With Nomogram Prediction Model

In our previous work, we built a nomogram model to predict the risk of PTE with the same training cohort through R statistical analysis [30]. The AUC of this nomogram prediction model was 0.859 (95% CI 0.826-0.891) and sensitivity and specificity were 0.867 and 0.738, respectively. Compared with the traditional nomogram prediction model, the AUC value of the ANN prediction model was higher (0.907, 95% CI 0.889-0.924; *P*=.01), which indicated the ANN model had a higher prediction accuracy.

Discussion

Principal Findings

ANN is applied to develop a risk prediction model and is superior to traditional prediction models. In this study, we developed a PTE predictive model using ANN methods, which involved 21 predictors (listed in Table 1). In the training cohort, the ANN model had an accuracy of 0.907, average precision of 0.557, sensitivity of 0.80, specificity of 0.86, PPV of 91%, and NPV of 86%. For testing cohort 1 (testing cohort 2), this model had an accuracy of 0.867 (0.859), average precision of 0.470 (0.344), sensitivity of 0.83 (0.80), specificity of 0.80 (0.84), PPV of 85% (78%), NPV of 80% (83%), and Brier score 0.121 (0.127), suggesting that this ANN model was valuable. To our knowledge, this is also the first study to develop a high-performance PTE-predictive ANN model based on other studies and risk factors available in clinical practice.

Advantages of the ANN Model

ANN models are able to model nonlinear relationships between input and output variables in a high-dimensional dataset and select the optimal model with high accuracy. ANN models have been widely used to predict the occurrence of hypertension [34] and mortality in patients with stroke [35]. ANN models excelled in many ways compared to conventional statistical methods; for example, they have higher classification accuracy and a better ability to analyze nonlinear relations and handle correlated independent variables [26].

Existing PTE prediction models are mainly risk scoring models built by traditional statistical methods. In our previous work, we set up a nomogram model to predict the risk of PTE. This model consisted of 7 risk factors (sex, TBI severity, duration of loss of consciousness time, subdural hematoma, early PTS, contusion site, and treatment) found in multivariable Cox proportional hazards regression analysis based on the same training cohort (West China Hospital). Our results showed that the AUC of this nomogram prediction model was 0.859 (95% CI 0.826-0.891), lower than the ANN model. In addition, with multivariable logistic regression and based on 9 significant risk factors (subdural hematoma, contusion load, craniotomy, craniectomy, seizure during acute hospitalization, duration of posttraumatic amnesia, preinjury mental health treatment/psychiatric hospitalization, intraparenchymal fragment, and preinjury incarceration), Ritter et al [36] constructed prognostic models to predict PTS during different times following TBI. Their results indicated that the corrected concordance statistics (equal to AUC) were 0.599, 0.747, and 0.716 for acute hospitalization, year 1, and year 2 models, respectively. In our study, we established an ANN model for PTE prediction using comprehensive data from training and 2 sets of tests that achieved AUC of 0.907, 0.867, and 0.859, respectively, higher than the existing models. In addition, Ritter et al [36] tested their model against subjects selected in bootstrap samples, while our ANN model was tested by the other 2 cohorts who were unaware of the training process. Our prediction model outperforms the abovementioned predictive models built by logistic regression method, which suggests that the ANN models have superiority and rationality in solving complex nonlinear relationships.

The new ANN model based on demographic and clinical data can be used as a simple screening tool to identify individuals at high risk of PTE after TBI. The predictors included in the model are common and available in routine practice. Beyond that, this model was tested by 2 cohorts and its performance was good, indicating that it might be applicable to the general population. In our study, we input 21 variables into the ANN model to predict the risk of PTE; all of these variables were mentioned in previous studies, while only some factors were considered as predictors of PTE by logistic regression. The ANN method has the advantage of feature selection over conventional statistical methods; when more factors are taken into account, the prediction is more accurate.

Impact of the ANN Model in the Future

The ANN model has higher prediction accuracy and can contribute to future clinical decisions. It helps clinicians identify patients with high risk of PTE, so doctors follow them more closely after discharge and follow up more frequently for more precise personal management. Furthermore, the new model is also conducive to the selection of the target group for PTE prevention study. For example, by applying presumed data, a provider could estimate a TBI patient's risk of PTE in the future. By studying the high-risk population predicted by the ANN model, it may be easier to find useful preventive measures. In addition, the ANN model can help clinicians conduct some trials on antiepileptic drug withdrawal. If according to the ANN model, the patient's PTE risk is low and meets the withdrawal criteria, the clinician may try to withdraw the patient's antiepileptic drug, which will reduce the financial burden and adverse effects of antiepileptic drugs.

Limitations

However, there were some limitations to this study. First, we developed the ANN model using epidemiological data, mainly including demographic data, clinical manifestation, and radiological results, regardless of relevant laboratory data such as electroencephalogram. Second, this was a retrospective study that was prone to bias, and some of the factors in other studies have not been collected, such as whether a patient was mentally ill. Third, most of the factors included were dichotomous variables rather than continuous variables. The lack of dose-response relationship between exposure levels of these risk factors and PTE may not reveal their true relationships with PTE. Fourth, the ANN model relied more on computers and specific programs, so its application was not as convenient and simple as nomogram models for the clinicians [37].

Despite these shortcomings, as far as we know, this is the first study using ANN to predict the risk of PTE after TBI. Our results indicated that ANN analysis may be more accurate in predicting the incidence of PTE for individual patients than traditional statistical methods, and therefore the ANN model could help determine the use of antiepileptic drugs for individual TBI patients.

Conclusions

In conclusion, our study was the first one to develop an ANN with a higher level of accurate prediction of PTE than the nomogram prediction model and other models constructed by multilogical regression. With the new ANN model, we can identify TBI patients at high risk of PTE as early as possible, and the model-predicted risk probability is significant for the selection of study population to determine the beneficial prevention and management of these PTE patients.

Acknowledgments

This study was funded by grant 2019HXFH048 from the 1·3·5 Project for Disciplines of Excellence–Clinical Research Incubation Project, West China Hospital, Sichuan University, and grant 2021YFS0174 from the Science & Technology Department of Sichuan Province.

Authors' Contributions

LL had full access to the data and takes responsibility for the integrity of the data and the accuracy of the data analysis. LL, JZ, TL, and XW were responsible for the concept and design of the study. XW, HW, SC, LZ, and DC were responsible for acquisition, analysis, and interpretation of the data. DC and XW performed model training and testing. XW drafted the manuscript. DC tested and trained the ANN model.

Conflicts of Interest

None declared.

References

1. Pickelsimer EE, Selassie AW, Gu JK, Langlois JA. A population-based outcomes study of persons hospitalized with traumatic brain injury: operations of the South Carolina Traumatic Brain Injury Follow-up Registry. J Head Trauma Rehabil 2006;21(6):491-504. [doi: 10.1097/00001199-200611000-00004] [Medline: 17122680]
2. Hauser WA, Annegers JF, Kurland LT. Incidence of epilepsy and unprovoked seizures in Rochester, Minnesota: 1935-1984. Epilepsia 1993;34(3):453-468. [doi: 10.1111/j.1528-1157.1993.tb02586.x] [Medline: 8504780]
3. Annegers JF, Hauser WA, Coan SP, Rocca WA. A population-based study of seizures after traumatic brain injuries. N Engl J Med 1998 Jan 01;338(1):20-24. [doi: 10.1056/NEJM199801013380104] [Medline: 9414327]
4. Englander J, Bushnik T, Duong TT, Cifu DX, Zafonte R, Wright J, et al. Analyzing risk factors for late posttraumatic seizures: a prospective, multicenter investigation. Arch Phys Med Rehabil 2003 Mar;84(3):365-373. [doi: 10.1053/apmr.2003.50022] [Medline: 12638104]
5. Christensen J, Pedersen MG, Pedersen CB, Sidenius P, Olsen J, Vestergaard M. Long-term risk of epilepsy after traumatic brain injury in children and young adults: a population-based cohort study. Lancet 2009 Mar 28;373(9669):1105-1110. [doi: 10.1016/S0140-6736(09)60214-2] [Medline: 19233461]
6. Angeleri F, Majkowski J, Cacchiò G, Sobieszek A, D'Acunto S, Gesuita R, et al. Posttraumatic epilepsy risk factors: one-year prospective study after head injury. Epilepsia 1999 Sep;40(9):1222-1230 [FREE Full text] [doi: 10.1111/j.1528-1157.1999.tb00850.x] [Medline: 10487184]
7. Asikainen I, Kaste M, Sarna S. Early and late posttraumatic seizures in traumatic brain injury rehabilitation patients: brain injury factors causing late seizures and influence of seizures on long-term outcome. Epilepsia 1999 May;40(5):584-589 [FREE Full text] [doi: 10.1111/j.1528-1157.1999.tb05560.x] [Medline: 10386527]
8. Armstrong KK, Sahgal V, Bloch R, Armstrong KJ, Heinemann A. Rehabilitation outcomes in patients with posttraumatic epilepsy. Arch Phys Med Rehabil 1990 Feb;71(2):156-160. [Medline: 2105709]
9. De Santis A, Sganzerla E, Spagnoli D, Bello L, Tiberio F. Risk factors for late posttraumatic epilepsy. Acta neurochirurgica Supplementum 1992;55:64-67. [doi: 10.1007/978-3-7091-9233-7_17] [Medline: 1414548]
10. Haltiner AM, Temkin NR, Dikmen SS. Risk of seizure recurrence after the first late posttraumatic seizure. Arch Phys Med Rehabil 1997 Aug;78(8):835-840. [doi: 10.1016/s0003-9993(97)90196-9] [Medline: 9344302]
11. Salazar AM, Schwab K, Grafman JH. Penetrating injuries in the Vietnam war. Traumatic unconsciousness, epilepsy, and psychosocial outcome. Neurosurg Clin N Am 1995 Oct;6(4):715-726. [Medline: 8527913]
12. Weiss GH, Feeney DM, Caveness WF, Dillon D, Kistler JP, Mohr JP, et al. Prognostic factors for the occurrence of posttraumatic epilepsy. Arch Neurol 1983 Jan;40(1):7-10. [doi: 10.1001/archneur.1983.04050010027006] [Medline: 6848098]
13. Rao VR, Parko KL. Clinical approach to posttraumatic epilepsy. Semin Neurol 2015 Feb;35(1):57-63. [doi: 10.1055/s-0035-1544239] [Medline: 25714868]
14. Yeh C, Chen T, Hu C, Chiu W, Liao C. Risk of epilepsy after traumatic brain injury: a retrospective population-based cohort study. J Neurol Neurosurg Psychiatry 2013 Apr;84(4):441-445. [doi: 10.1136/jnnp-2012-302547] [Medline: 23117492]
15. Emanuelson I, Uvebrant P. Occurrence of epilepsy during the first 10 years after traumatic brain injury acquired in childhood up to the age of 18 years in the south western Swedish population-based series. Brain Inj 2009 Jul;23(7):612-616. [doi: 10.1080/02699050902973913] [Medline: 19557563]
16. Tubi MA, Lutkenhoff E, Blanco MB, McArthur D, Villablanca P, Ellingson B, et al. Early seizures and temporal lobe trauma predict post-traumatic epilepsy: a longitudinal study. Neurobiol Dis 2019 Mar;123:115-121 [FREE Full text] [doi: 10.1016/j.nbd.2018.05.014] [Medline: 29859872]
17. Zhao Y, Wu H, Wang X, Li J, Zhang S. Clinical epidemiology of posttraumatic epilepsy in a group of Chinese patients. Seizure 2012 Jun;21(5):322-326 [FREE Full text] [doi: 10.1016/j.seizure.2012.02.007] [Medline: 22425008]

18. Kruer RM, Harris LH, Goodwin H, Kornbluth J, Thomas KP, Slater LA, et al. Changing trends in the use of seizure prophylaxis after traumatic brain injury: a shift from phenytoin to levetiracetam. J Crit Care 2013 Oct;28(5):883-913. [doi: 10.1016/j.jcrc.2012.11.020] [Medline: 23566730]
19. Thompson K, Pohlmann-Eden B, Campbell LA, Abel H. Pharmacological treatments for preventing epilepsy following traumatic head injury. Cochrane Database Syst Rev 2015 Aug 10(8):CD009900. [doi: 10.1002/14651858.CD009900.pub2] [Medline: 26259048]
20. Agarwal V, Zhang L, Zhu J, Fang S, Cheng T, Hong C, et al. Impact of predicting health care utilization via web search behavior: a data-driven analysis. J Med Internet Res 2016 Sep 21;18(9):e251 [FREE Full text] [doi: 10.2196/jmir.6240] [Medline: 27655225]
21. Manning T, Sleator RD, Walsh P. Biologically inspired intelligent decision making: a commentary on the use of artificial neural networks in bioinformatics. Bioengineered 2014;5(2):80-95 [FREE Full text] [doi: 10.4161/bioe.26997] [Medline: 24335433]
22. 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]
23. Tu JV. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. J Clin Epidemiol 1996 Nov;49(11):1225-1231. [doi: 10.1016/s0895-4356(96)00002-9] [Medline: 8892489]
24. Chan KL, Leng X, Zhang W, Dong W, Qiu Q, Yang J, et al. Early identification of high-risk TIA or minor stroke using artificial neural network. Front Neurol 2019;10:171 [FREE Full text] [doi: 10.3389/fneur.2019.00171] [Medline: 30881336]
25. Santus E, Marino N, Cirillo D, Chersoni E, Montagud A, Santuccione Chadha A, et al. Artificial intelligence-aided precision medicine for COVID-19: strategic areas of research and development. J Med Internet Res 2021 Mar 12;23(3):e22453 [FREE Full text] [doi: 10.2196/22453] [Medline: 33560998]
26. Renganathan V. Overview of artificial neural network models in the biomedical domain. Bratisl Lek Listy 2019;120(7):536-540. [doi: 10.4149/BLL_2019_087] [Medline: 31602991]
27. Fisher RS, Acevedo C, Arzimanoglou A, Bogacz A, Cross JH, Elger CE, et al. ILAE official report: a practical clinical definition of epilepsy. Epilepsia 2014 Apr;55(4):475-482 [FREE Full text] [doi: 10.1111/epi.12550] [Medline: 24730690]
28. Bentley P, Ganesalingam J, Carlton Jones AL, Mahady K, Epton S, Rinne P, et al. Prediction of stroke thrombolysis outcome using CT brain machine learning. Neuroimage Clin 2014;4:635-640 [FREE Full text] [doi: 10.1016/j.nicl.2014.02.003] [Medline: 24936414]
29. Zhou Z, Folkert M, Cannon N, Iyengar P, Westover K, Zhang Y, et al. Predicting distant failure in early stage NSCLC treated with SBRT using clinical parameters. Radiother Oncol 2016 Jun;119(3):501-504 [FREE Full text] [doi: 10.1016/j.radonc.2016.04.029] [Medline: 27156652]
30. Wang X, Zhong J, Lei T, Wang H, Zhu L, Chu S, et al. Development and external validation of a predictive nomogram model of posttraumatic epilepsy: a retrospective analysis. Seizure 2021 May;88:36-44. [doi: 10.1016/j.seizure.2021.03.023] [Medline: 33812306]
31. Davis J, Goadrich M. The relationship between precision-recall and ROC curves. Proc 23rd Int Conf Mach Learn 2006:233-240. [doi: 10.1145/1143844.1143874]
32. Brier GW. Verification of forecasts expressed in terms of probability. Mon Wea Rev 1950 Jan;78(1):1-3. [doi: 10.1175/1520-0493(1950)078<0001:vofeit>2.0.co;2]
33. Steyerberg EW, Harrell FE, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. J Clin Epidemiol 2001 Aug;54(8):774-781. [doi: 10.1016/s0895-4356(01)00341-9] [Medline: 11470385]
34. Huang S, Xu Y, Yue L, Wei S, Liu L, Gan X, et al. Evaluating the risk of hypertension using an artificial neural network method in rural residents over the age of 35 years in a Chinese area. Hypertens Res 2010 Jul;33(7):722-726. [doi: 10.1038/hr.2010.73] [Medline: 20505678]
35. Çelik G, Baykan, Kara Y, Tireli H. Predicting 10-day mortality in patients with strokes using neural networks and multivariate statistical methods. J Stroke Cerebrovasc Dis 2014 Jul;23(6):1506-1512. [doi: 10.1016/j.jstrokecerebrovasdis.2013.12.018] [Medline: 24674954]
36. Ritter AC, Wagner AK, Szaflarski JP, Brooks MM, Zafonte RD, Pugh MJV, et al. Prognostic models for predicting posttraumatic seizures during acute hospitalization, and at 1 and 2 years following traumatic brain injury. Epilepsia 2016 Sep;57(9):1503-1514 [FREE Full text] [doi: 10.1111/epi.13470] [Medline: 27430564]
37. Gotto GT, Yu C, Bernstein M, Eastham JA, Kattan MW. Development of a nomogram model predicting current bone scan positivity in patients treated with androgen-deprivation therapy for prostate cancer. Front Oncol 2014;4:296 [FREE Full text] [doi: 10.3389/fonc.2014.00296] [Medline: 25386410]

Abbreviations

ANN: artificial neural network
AUC: area under the curve
ICD-10: International Statistical Classification of Diseases and Related Health Problems, Tenth Revision

NPV: negative predictive value
PPV: positive predictive value
PTE: posttraumatic epilepsy
PTS: posttraumatic seizure
ROC: receiver operating characteristic curve
TBI: traumatic brain injury

Edited by R Kukafka; submitted 16.10.20; peer-reviewed by L Zhang, M Sotiropoulos; comments to author 26.02.21; revised version received 14.04.21; accepted 25.04.21; published 19.08.21.

Please cite as:
Wang X, Zhong J, Lei T, Chen D, Wang H, Zhu L, Chu S, Liu L
An Artificial Neural Network Prediction Model for Posttraumatic Epilepsy: Retrospective Cohort Study
J Med Internet Res 2021;23(8):e25090
URL: https://www.jmir.org/2021/8/e25090
doi:10.2196/25090
PMID:34420931

©Xueping Wang, Jie Zhong, Ting Lei, Deng Chen, Haijiao Wang, Lina Zhu, Shanshan Chu, Ling Liu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Review

Computer-Aided Diagnosis of Diminutive Colorectal Polyps in Endoscopic Images: Systematic Review and Meta-analysis of Diagnostic Test Accuracy

Chang Seok Bang[1,2,3,4], MD, PhD; Jae Jun Lee[3,4,5], MD; Gwang Ho Baik[1,2], MD

[1]Department of Internal Medicine, Hallym University College of Medicine, Chuncheon, Republic of Korea
[2]Institute for Liver and Digestive Diseases, Hallym University College of Medicine, Chuncheon, Republic of Korea
[3]Institute of New Frontier Research, Hallym University College of Medicine, Chuncheon, Republic of Korea
[4]Division of Big Data and Artificial Intelligence, Chuncheon Sacred Heart Hospital, Chuncheon, Republic of Korea
[5]Department of Anesthesiology and Pain Medicine, Hallym University College of Medicine, Chuncheon, Republic of Korea

Corresponding Author:
Chang Seok Bang, MD, PhD
Department of Internal Medicine
Hallym University College of Medicine
Sakju-ro 77
Chuncheon, 24253
Republic of Korea
Phone: 82 332405821
Fax: 82 332418064
Email: csbang@hallym.ac.kr

Abstract

Background: Most colorectal polyps are diminutive and benign, especially those in the rectosigmoid colon, and the resection of these polyps is not cost-effective. Advancements in image-enhanced endoscopy have improved the optical prediction of colorectal polyp histology. However, subjective interpretability and inter- and intraobserver variability prohibits widespread implementation. The number of studies on computer-aided diagnosis (CAD) is increasing; however, their small sample sizes limit statistical significance.

Objective: This review aims to evaluate the diagnostic test accuracy of CAD models in predicting the histology of diminutive colorectal polyps by using endoscopic images.

Methods: Core databases were searched for studies that were based on endoscopic imaging, used CAD models for the histologic diagnosis of diminutive colorectal polyps, and presented data on diagnostic performance. A systematic review and diagnostic test accuracy meta-analysis were performed.

Results: Overall, 13 studies were included. The pooled area under the curve, sensitivity, specificity, and diagnostic odds ratio of CAD models for the diagnosis of diminutive colorectal polyps (adenomatous or neoplastic vs nonadenomatous or nonneoplastic) were 0.96 (95% CI 0.93-0.97), 0.93 (95% CI 0.91-0.95), 0.87 (95% CI 0.76-0.93), and 87 (95% CI 38-201), respectively. The meta-regression analysis showed no heterogeneity, and no publication bias was detected. Subgroup analyses showed robust results. The negative predictive value of CAD models for the diagnosis of adenomatous polyps in the rectosigmoid colon was 0.96 (95% CI 0.95-0.97), and this value exceeded the threshold of the *diagnosis and leave* strategy.

Conclusions: CAD models show potential for the optical histological diagnosis of diminutive colorectal polyps via the use of endoscopic images.

Trial Registration: PROSPERO CRD42021232189; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=232189

(J Med Internet Res 2021;23(8):e29682) doi:10.2196/29682

KEYWORDS

artificial intelligence; deep learning; polyps; colon; colonoscopy; diminutive

Introduction

Colorectal cancer (CRC) is the third most common cancer based on incidence statistics and the second leading cause of cancer-related deaths worldwide [1]. Most CRCs arise from benign neoplastic polyps, such as adenomas [2]. Colonoscopy with the identification and removal of these neoplastic polyps is a standard screening method for CRC, which has been proven to reduce cancer-related mortality [2,3]. This is because polyp removal through colonoscopy prevents the development of CRCs by interrupting the adenoma-carcinoma sequence, which is the most reliable stepwise pathogenesis of CRC development [4].

With regard to the size of colorectal polyps, 90%-95% of the detected polyps are <1 cm, and about half of them are nonneoplastic [3-6]. In the context of diminutive colorectal polyps (DCPs; ≤5 mm), only 0.5%-1.7% of cases had advanced histology, indicating a lower probability of developing CRCs [4,7-9]. However, current practice points out the removal of all detected polyps and sending them for histologic evaluation [10]. This may help determine the surveillance interval for CRC screening with per-patient risk stratification [8]. However, unnecessary polypectomy carries the risk of procedure-related adverse events and is not cost-effective [4,8].

With the advancement of image-enhanced endoscopy, optical diagnosis has been attempted to predict the histology of the detected polyps during colonoscopy by characterizing the surface morphology. This can reduce the need for histologic evaluation after the removal of neoplastic lesions with a small risk of having an invasive component [11]. The unnecessary removal of benign polyps can be avoided with the adaptation of this technique. Therefore, optical diagnosis using electronic or dye-based methods has been recommended for histological classification in clinical practice [10]. In accordance with this technique, Preservation and Incorporation of Valuable Endoscopic Innovations (PIVI) performance thresholds for in situ endoscopic histology prediction (optical biopsy) required for *resect and discard* and *diagnose and leave* strategies have been suggested for the management of DCPs [12]. For the management of polyps suspected as neoplasm with diminutive size based on the optical biopsy, a *resect and discard* strategy should satisfy >90% agreement in postpolypectomy surveillance intervals compared with histologic assessment [4]. For nonneoplastic polyps <5 mm in the rectosigmoid colon, the negative predictive value (NPV) should be >90% to adopt the *diagnose and leave* strategy based on the optical biopsy in PIVI performance thresholds [4]. However, only studies by experienced endoscopists with a high level of confidence showed benefits in optical biopsy [10]. Subjective interpretability, inter- or intraobserver variability, and the learning curve prohibits the widespread implementation of this technique.

Studies on computer-aided diagnosis (CAD) using deep learning or machine learning methods to define the accuracy of CAD models are increasing [13,14]. The performance of the CAD model was not influenced by the endoscopists' level of confidence, and the CAD model consistently provided robust answers. However, studies with small sample sizes have inadequate statistical strength. Thus, this study aims to evaluate the diagnostic test accuracy (DTA) of CAD models used for the histologic diagnosis of DCPs using endoscopic images.

Methods

Adherence to the Statement of Systematic Review and Protocol Administration

This study was conducted in accordance with the PRISMA (Preferred Reporting Items for Systematic Review and Meta-Analyses) of DTA Studies [15]. The study protocol was registered at the International Prospective Register of Systematic Reviews database before the initiation of the systematic review (CRD42021232189). Approval from the institutional review board of the Chuncheon Sacred Heart Hospital was waived.

Literature Search

Two authors (CSB and JJL) independently performed a core database search of MEDLINE, PubMed, Embase, and Cochrane Library using common search formulas, from inception to January 2020. Duplicate articles were excluded from the analyses. The titles and abstracts of all identified articles were reviewed, and irrelevant articles were excluded. Full-text reviews were subsequently performed to determine whether the pre-established inclusion criteria were satisfied in the identified literature. References were also reviewed to identify any additional relevant articles. Any disagreements in the results of the search process between the 2 authors were resolved by discussion or consultation with a third author (GHB). The search formulas used to identify the relevant articles are presented in Textbox 1.

Textbox 1. Literature search strategy for the core databases. tiab: searching code for title and abstract; Mesh: Medical Subject Headings; ab,ti,kw: searching code for abstract, title, and keywords; Lang: searching code for language; lim: searching code by limiting certain conditions.

MEDLINE (Through PubMed)

1. *artificial intelligence*[tiab] OR *AI*[tiab] OR *deep learning*[tiab] OR *machine learning*[tiab] OR *computer*[tiab] OR *neural network*[tiab] OR *CNN*[tiab] OR *automatic*[tiab] OR *automated*[tiab]: 502318
2. *diminutive*[tiab] OR *small*[tiab]: 1417047
3. *polyp*[tiab] OR *polyps*[Mesh]: 40395
4. 1 AND 2 AND 3: 128
5. 5-4 AND English[Lang]: 125

Embase

1. *artificial intelligence*:ab,ti,kw OR AI:ab,ti,kw OR *deep learning*:ab,ti,kw OR *machine learning*:ab,ti,kw OR *computer*:ab,ti,kw OR *neural network*:ab,ti,kw OR *CNN*:ab,ti,kw OR *automatic*:ab,ti,kw OR *automated*: 638513
2. *diminutive*:ab,ti,kw OR *small*: ab,ti,kw: 2056711
3. *polyp*:ab,ti,kw: 29836
4. 3-1 AND 2 AND 3: 198
5. 4-3 AND ([article]/lim OR [article in press]/lim OR [review]/lim) AND [English]/lim: 104

Cochrane Library

1. artificial intelligence:ab,ti,kw or AI:ab,ti,kw or deep learning:ab,ti,kw or machine learning:ab,ti,kw or computer:ab,ti,kw or neural network:ab,ti,kw or CNN:ab,ti,kw or automatic:ab,ti,kw or automated:ab,ti,kw: 56749
2. Mesh descriptor: [polyps] explode all trees: 1087
3. polyp:ab,ti,kw: 2855
4. 2 or 3: 3397
5. diminutive:ab,ti,kw or small:ab,ti,kw: 83388
6. 1 and 4 and 5: 48 trials (2021-1-28)

Literature Selection Criteria

The literature included in this systematic review should meet the following inclusion criteria: designed to evaluate the diagnostic performance of CAD models in the prediction of histology of DCPs based on endoscopic images; presentation of the diagnostic performance of CAD models, including sensitivity, specificity, likelihood ratios, predictive values, or accuracy, which enabled the estimation of true-positive (TP), false-positive (FP), false-negative (FN), and true-negative (TN) values for the histologic diagnosis of DCPs based on endoscopic images; and studies written in English. The exclusion criteria were as follows: narrative review articles; studies with incomplete data; systematic reviews or meta-analyses; and comments, proceedings, or study protocols. Articles meeting at least one of the exclusion criteria were excluded from this systematic review.

Methodological Quality Evaluation

Two authors (CSB and JJL) assessed the methodological quality of the final included articles using the second version of the Quality Assessment of Diagnostic Accuracy Studies. This tool comprises four domains: *patient selection*, *index test*, *reference standard*, and *flow and timing*, and the first three domains have an applicability assessment. The 2 authors assessed each part as having a high, low, or unclear risk of bias [16].

Data Extraction, Primary Outcomes, and Additional Analyses

Two authors (CSB and JJL) independently extracted the data from each included study and cross-checked the extracted data. If the data were unclear, the corresponding author of the study was contacted by email to obtain insight into the original data set. A descriptive synthesis was performed using a systematic review process, and DTA meta-analysis was conducted if the included studies were sufficiently homogenous.

The primary outcomes were the TP, FP, FN, and TN values in each study. For the CAD of the histology of DCPs using endoscopic images, the primary outcomes were defined as follows: TP referred to the number of subjects with a positive finding by a CAD model and have adenomas or neoplasms as evidenced by endoscopic images; FP referred to the number of subjects with a positive finding by a CAD model and do not have adenomas or neoplasms based on endoscopic images; FN referred to the number of subjects with a negative finding by a CAD model and have adenomas or neoplasms as evidenced by endoscopic images; and TN referred to the number of subjects with a negative finding on a CAD model and do not have adenomas or neoplasms based on the endoscopic images. With these definitions, the TP, FP, FN, and TN values were calculated for each included study. If the included studies presented comparative diagnostic performance of endoscopists versus

CAD models, the TP, FP, FN, and TN values of endoscopists in each study were also extracted.

For additional analyses, such as subgroup analysis or meta-regression, the authors extracted the following variables from each included study: publication year, geographic origin of the data (ie, Western vs Asian), type of endoscopic images, type of CAD models, location of the DCPs (ie, any colon vs rectosigmoid colon), number of total images included, and type of test data sets (internal test vs external test).

Statistics

The bivariate method [17] and hierarchical summary receiver operating characteristic (HSROC) method [18] were applied for the DTA meta-analysis. A forest plot of the sensitivity and specificity and a summary receiver operating characteristic (SROC) curve were generated using the bivariate method [17] and HSROC [18] method, respectively. The level of heterogeneity across the included articles was determined by the correlation coefficient between logit-transformed sensitivity and specificity by the bivariate method and the asymmetry parameter β, where $\beta=0$ corresponds to a symmetric receiver operating characteristic curve, in which the diagnostic odds ratio (DOR) does not vary along the curve according to the HSROC method. A positive correlation coefficient and a β with a significant probability ($P<.05$) indicated heterogeneity between the studies [18,19]. A visual examination of the SROC curve was also performed to identify heterogeneity. Subgroup analysis by univariate meta-regression using the modifiers identified during the systematic review was also performed to identify the reasons for heterogeneity. The pooled NPV by integrating conditional prevalence with respect to a previous distribution (considering the heterogeneity in prevalence) was calculated using a probability-modifying plot.

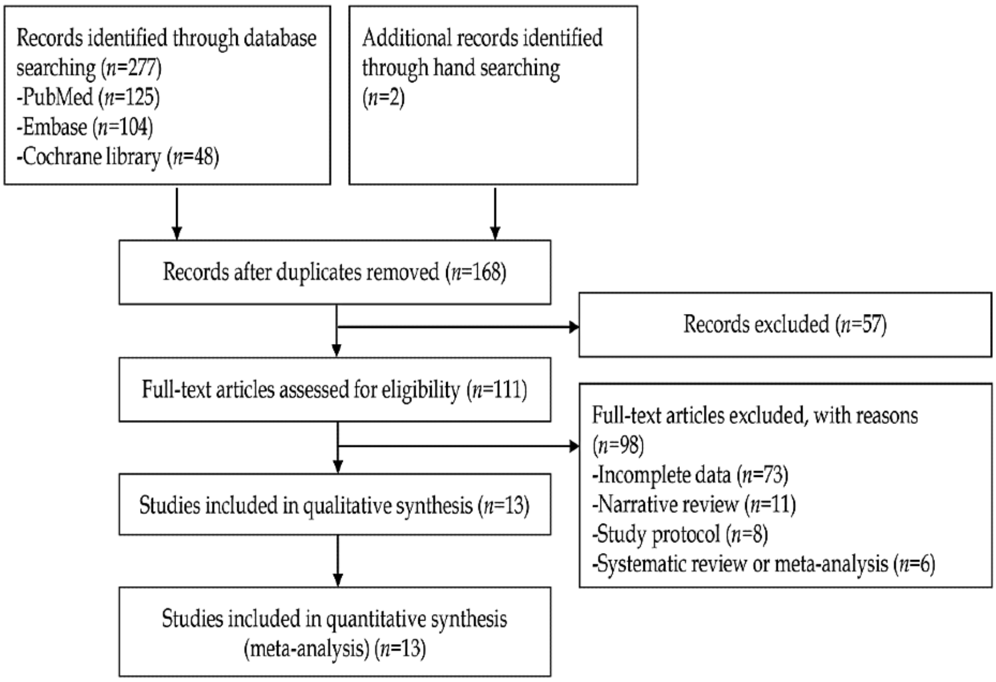
STATA software version 15.1, including the METANDI and MIDAS packages, was used for the DTA meta-analysis. The METANDI and MIDAS packages require the inclusion of a minimum of four studies for DTA meta-analysis. Therefore, if less than four studies were included in the subgroup analysis, the Moses-Shapiro-Littenberg method [20], as implemented in Meta-DiSc 1.4 (XI Cochrane Colloquium), was used. Publication bias was evaluated using the Deek funnel plot asymmetry test.

Results

Study Selection

A total of 277 articles were identified following a literature search of the three core databases. Two additional studies were identified by manual screening of the bibliographies. After excluding 111 duplicate studies, 57 additional articles were excluded after reviewing the titles and abstracts. Full-text versions of the remaining 111 articles were obtained and thoroughly reviewed based on the aforementioned inclusion and exclusion criteria. Among these, 98 articles were excluded from the final enrollment for the following reasons: 73 (74%) for incomplete data, 11 (11%) for narrative review, 8 (8%) for study protocol, and 6 (6%) for systematic review or meta-analysis. Finally, 13 studies [21-33] were included in the systematic review. A flowchart of the selection process is presented in Figure 1.

Figure 1. Flowchart of the selection process.


Records identified through database searching (n=277)
-PubMed (n=125)
-Embase (n=104)
-Cochrane library (n=48)
Additional records identified through hand searching (n=2)
Records after duplicates removed (n=168)
Records excluded (n=57)
Full-text articles assessed for eligibility (n=111)
Full-text articles excluded, with reasons (n=98)
-Incomplete data (n=73)
-Narrative review (n=11)
-Study protocol (n=8)
-Systematic review or meta-analysis (n=6)
Studies included in qualitative synthesis (n=13)
Studies included in quantitative synthesis (meta-analysis) (n=13)

Clinical Features in Included Studies

The identified studies established and explored the diagnostic performance of CAD models for the classification of adenomatous or neoplastic versus nonadenomatous or nonneoplastic polyps. Among the 13 studies for the CAD of DCPs, 6564 images were identified (3207 cases vs 3357 controls).

Seven studies [24,26,27,29-32] used endoscopic images from Asian populations, and six studies [21-23,25,28,33] used endoscopic images from Western populations. All included studies adopted the definition of DCPs as size <5 mm. However, Shahidi et al [23] adopted a stricter definition for DCPs with a size <3 mm. With regard to the type of CAD model, a deep neural network or convolutional neural network was used in six studies [21-25,30], a support vector machine in six studies [27-29,31-33], and a software-based automatic color intensity analysis in one study [26]. White-light imaging is currently the standard method for inspecting endoscopic lesions. However, one study [28] used white-light imaging to establish a CAD model, and most of the included studies used image-enhanced endoscopic images, such as narrow-band imaging [21-26,30,32,33] or autofluorescence imaging [26] with or without magnification for the detailed characterization of the morphology of DCPs. Three studies [27,29,31] have used images of endocytoscopy, which is a specialized endoscopy that allows the analysis of mucosal structures at the cellular level [34]. With regard to the location of DCPs, most studies [21,23-25,27,28,30-33] did not consider the location of DCPs. However, three studies [22,26,29] separately collected DCPs from the rectosigmoid colon and evaluated the diagnostic performance of the CAD model for these polyps. Most of the included studies [21,23-26,28-33], except for two studies [22,27], have evaluated diagnostic performance using an internal test data set. A study by Zachariah et al [22] presented both external and internal test performance, and a study by Kudo et al [27] presented an external test format of the CAD model previously established by Mori et al in 2016 [31] and 2018 [29,31]. Five studies [25,27,30,31,33] have presented the comparative diagnostic performance of endoscopists versus CAD models for the prediction of histology in DCPs. Detailed clinical features of the included studies are presented in Table 1.

Table 1. Clinical characteristics of the included studies.

Study (year)	Nationality (data)	Definition of diminutive polyp	Type of CAD[a] models	Type of endoscopic images	Type of case and controls	Location of polyps	Type of test data sets	Number of cases in test data set (adenoma)	Number of controls in test data set	TP[b]	FP[c]	FN[d]	TN[e]	Performance of endoscopists (TP/FP/FN/TN)
Eladio Rodriguez-Diaz et al (2020) [21]	United States	≤5 mm	CNN[f]	NBI[g] with near focus magnification	Neoplastic versus nonneoplastic polyp	All	Internal test	93	75	88	9	5	66	N/A[h]
Zachariah et al (2020) [22]	United States	≤5 mm	CNN	WLI[i] or NBI	Adenomatous versus nonadenomatous	RS[j] colon	Internal test	119	472	107	38	12	434	N/A
Zachariah et al (2020) [22]	United States	≤5 mm	CNN	WLI or NBI	Adenomatous versus nonadenomatous	RS colon	External test	183	503	167	60	16	443	N/A
Shahidi et al (2020) [23]	Canada	≤3 mm	CNN	NBI with or without near focus magnification	Adenomatous versus nonadenomatous	All	Internal test	458	186	409	168	49	18	N/A
Jin et al (2020) [24]	South Korea	≤5 mm	CNN	NBI with or without near focus magnification	Adenomatous versus hyperplastic polyp	All	Internal test	180	120	150	10	30	110	N/A
Byrne et al (2019) [25]	Canada	≤5 mm	CNN	NBI with or without near focus magnification	Adenomatous versus hyperplastic polyp	All	Internal test	66	40	65	7	1	33	43/15/9/35
Horiuchi et al (2019) [26]	Japan	≤5 mm	Software-based automatic color intensity analysis	AFI[k]	Neoplastic versus nonneoplastic polyp	All	Internal test	212	217	164	15	48	202	N/A
Horiuchi et al (2019) [26]	Japan	≤5 mm	Software-based automatic color intensity analysis	TME[l] (WLI, NBI with magnification, and AFI)	Neoplastic versus nonneoplastic polyp	All	Internal test	212	217	191	18	21	199	N/A
Horiuchi et al (2019) [26]	Japan	≤5 mm	Software-based automatic color intensity analysis	AFI	Neoplastic versus nonneoplastic polyp	RS colon	Internal test	65	193	52	9	13	184	N/A

Study (year)	Nationality (data)	Definition of diminutive polyp	Type of CAD[a] models	Type of endoscopic images	Type of case and controls	Location of polyps	Type of test data sets	Number of cases in test data set (adenoma)	Number of controls in test data set	TP[b]	FP[c]	FN[d]	TN[e]	Performance of endoscopists (TP/FP/FN/TN)
Horiuchi et al (2019) [26]	Japan	≤5 mm	Software-based automatic color intensity analysis	TME (WLI, NBI with magnification, and AFI)	Neoplastic versus non-neoplastic polyp	RS colon	Internal test	65	193	55	8	10	185	N/A
Kudo et al (2019) [27]	Japan	≤5 mm	SVM[m]	Endocytoscope with NBI	Neoplastic versus non-neoplastic polyp	All	External test	1000	680	960	40	40	640	459/12/41/328 (expert); 578/97/422/583 (trainee)
Kudo et al (2019) [27]	Japan	≤5 mm	SVM	Endocytoscope with CE[n] (methylene blue)	Neoplastic versus non-neoplastic polyp	All	External test	1000	680	960	0	40	680	453/20/47/320 (expert); 690/236/310/444 (trainee)
Cristina Sánchez-Montes et al (2019) [28]	Spain	≤5 mm	SVM	WLI	Neoplastic versus non-neoplastic polyp	All	Internal test	50	50	43	6	7	44	N/A
Mori et al (2018) [29]	Japan	≤5 mm	SVM	Endocytoscope with NBI	Neoplastic versus non-neoplastic polyp	All	Internal test	287	185	268	16	19	159	N/A
Mori et al (2018) [29]	Japan	≤5 mm	SVM	Endocytoscope with CE (methylene blue)	Neoplastic versus non-neoplastic polyp	All	Internal test	287	185	263	17	24	158	N/A
Mori et al (2018) [29]	Japan	≤5 mm	SVM	Endocytoscope with NBI	Neoplastic versus non-neoplastic polyp	RS colon	Internal test	104	144	98	6	6	138	N/A
Mori et al (2018) [29]	Japan	≤5 mm	SVM	Endocytoscope with CE (methylene blue)	Neoplastic versus non-neoplastic polyp	RS colon	Internal test	104	144	96	11	8	133	N/A
Chen et al (2018) [30]	Taiwan	≤5 mm	Deep neural network	NBI with magnification	Neoplastic versus hyperplastic polyp	All	Internal test	188	96	181	21	7	75	367/55/9/137 (expert); 671/95/81/289 (novice)
Mori et al (2016) [31]	Japan	≤5 mm	SVM	Endocytoscope with WLI	Neoplastic versus non-neoplastic polyp	All	Internal test	19	36	18	2	1	34	248/16/25/128 (expert); 646/106/264/374 (nonexpert)
Kominami et al (2016) [32]	Japan	≤5 mm	SVM	NBI with magnification	Neoplastic versus non-neoplastic polyp	All	Internal test	43	45	40	3	3	42	N/A

Study (year)	Nationality (data)	Definition of diminutive polyp	Type of CAD[a] models	Type of endoscopic images	Type of case and controls	Location of polyps	Type of test data sets	Number of cases in test data set (adenoma)	Number of controls in test data set	TP[b]	FP[c]	FN[d]	TN[e]	Performance of endoscopists (TP/FP/FN/TN)
Gross et al (2011) [33]	Germany	≤5 mm	SVM	NBI with magnification	Adenomatous versus nonneoplastic polyp	All	Internal test	140	135	133	11	7	124	217/17/23/253 (expert); 188/26/52/244 (nonexpert)

[a] CAD: computer-aided diagnosis.

[b] TP: true-positive.

[c] FP: false-positive.

[d] FN: false-negative.

[e] TN: true-negative.

[f] CNN: convolutional neural network.

[g] NBI: narrow-band imaging.

[h] N/A: not applicable.

[i] WLI: white-light imaging.

[j] RS: rectosigmoid.

[k] AFI: autofluorescence imaging.

[l] TME: trimodal imaging endoscopy.

[m] SVM: support vector machine.

[n] CE: chromoendoscopy.

Quality Assessment of Study Methodology

The quality of the baseline image data is important because the CAD model is established using the learning features of the baseline training data. Theoretically, the images included in each study should reflect real-world conditions, as the CAD model was established for use in clinical practice. However, as some lesions are rare or abnormal, data imbalance is the main barrier to the learning of CAD models. Most of the included studies in the systematic review attempted to mitigate this pitfall by adopting specific inclusion and exclusion criteria for the enrollment of endoscopic images. However, four studies [21,23,25,28] did not include a detailed description of the image enrollment standard. Therefore, these studies were rated as *unclear risk* in the *patient selection* domain (Figures 2 and 3). This binary classification of *low risk* and *unclear risk* in the *patient selection* domain was adopted as a modifier in the subgroup or meta-regression analysis.

Figure 2. Summary graph of methodological quality. "+" denotes a low risk of bias, "?" denotes an unclear risk of bias, and "−" denotes a high risk of bias.

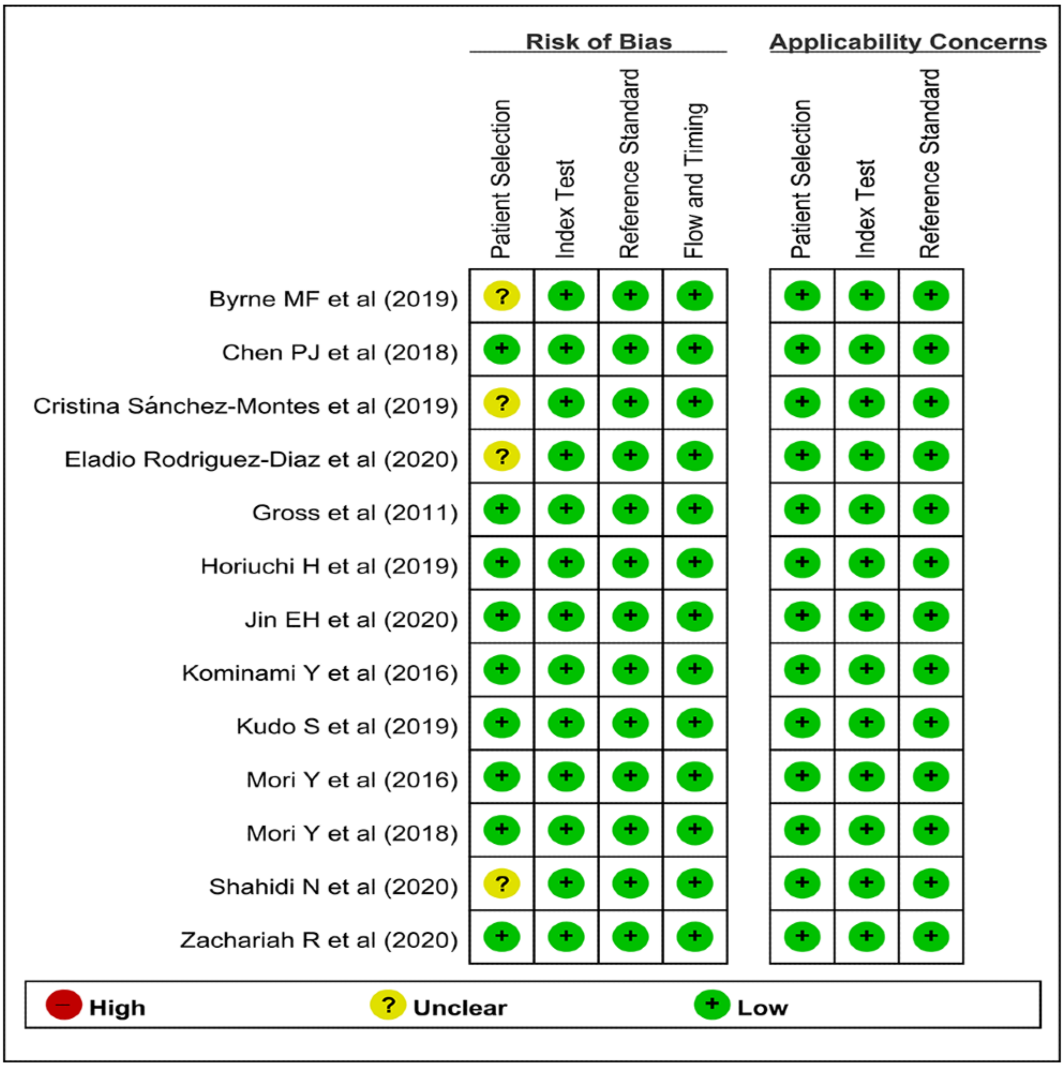
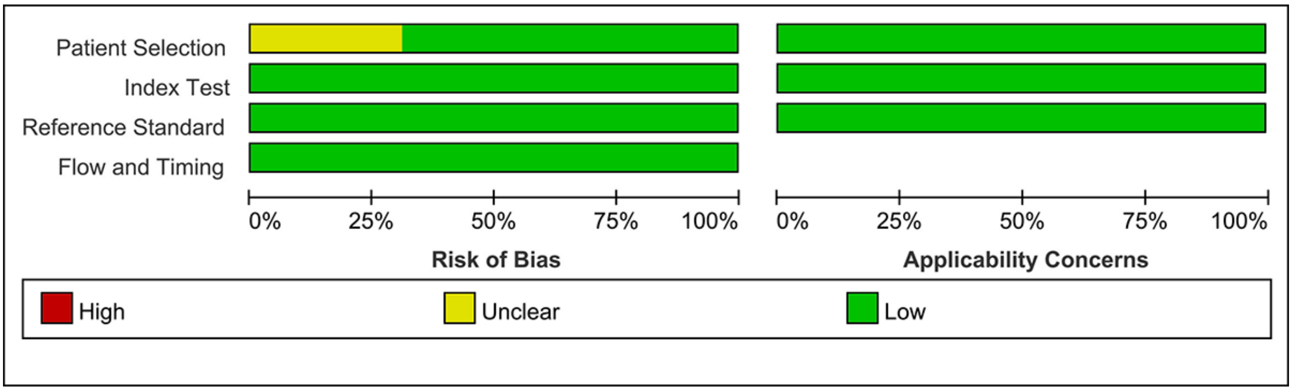


Figure 3. Summary table of methodological quality. "" denotes a low risk of bias, "?" denotes an unclear risk of bias, and "−" denotes a high risk of bias.



DTA Meta-analysis of CAD Models

Among the 13 studies [21-33] for the meta-analysis of CAD of DCPs, the area under the curve (AUC), sensitivity, specificity, positive likelihood ratio, negative likelihood ratio, and DOR of CAD models for the diagnosis of DCPs were 0.96 (95% CI 0.93-0.97), 0.93 (95% CI 0.91-0.95), 0.87 (95% CI 0.76-0.93), 7.1 (95% CI 3.8-13.3), 0.08 (95% CI 0.06-0.11), and 87 (95% CI 38-201), respectively (Figure 4; Table 2). The SROC curve is shown in Figure 5. To investigate the clinical utility of the CAD models, Fagan nomogram was generated. Positive findings indicated that adenomas or neoplasms were detected by the CAD models. Negative findings indicated that nonadenomas or nonneoplasms were detected by the CAD models. After assuming a 49% prevalence of adenomas or neoplasms (this value was calculated from the values in Table 1; ie, the total number of cases/controls, ie, 3207/3357, 95.53%), the Fagan nomogram shows that the posterior probability of adenomas or neoplasms was 87% if the finding of the CAD model was positive, and the posterior probability of adenoma was 7% if the finding of the CAD model was negative (Figure 6).

Figure 4. Coupled forest plots of sensitivity and specificity in computer-aided diagnosis models for the diagnosis of histology for diminutive colorectal polyps using endoscopic images.

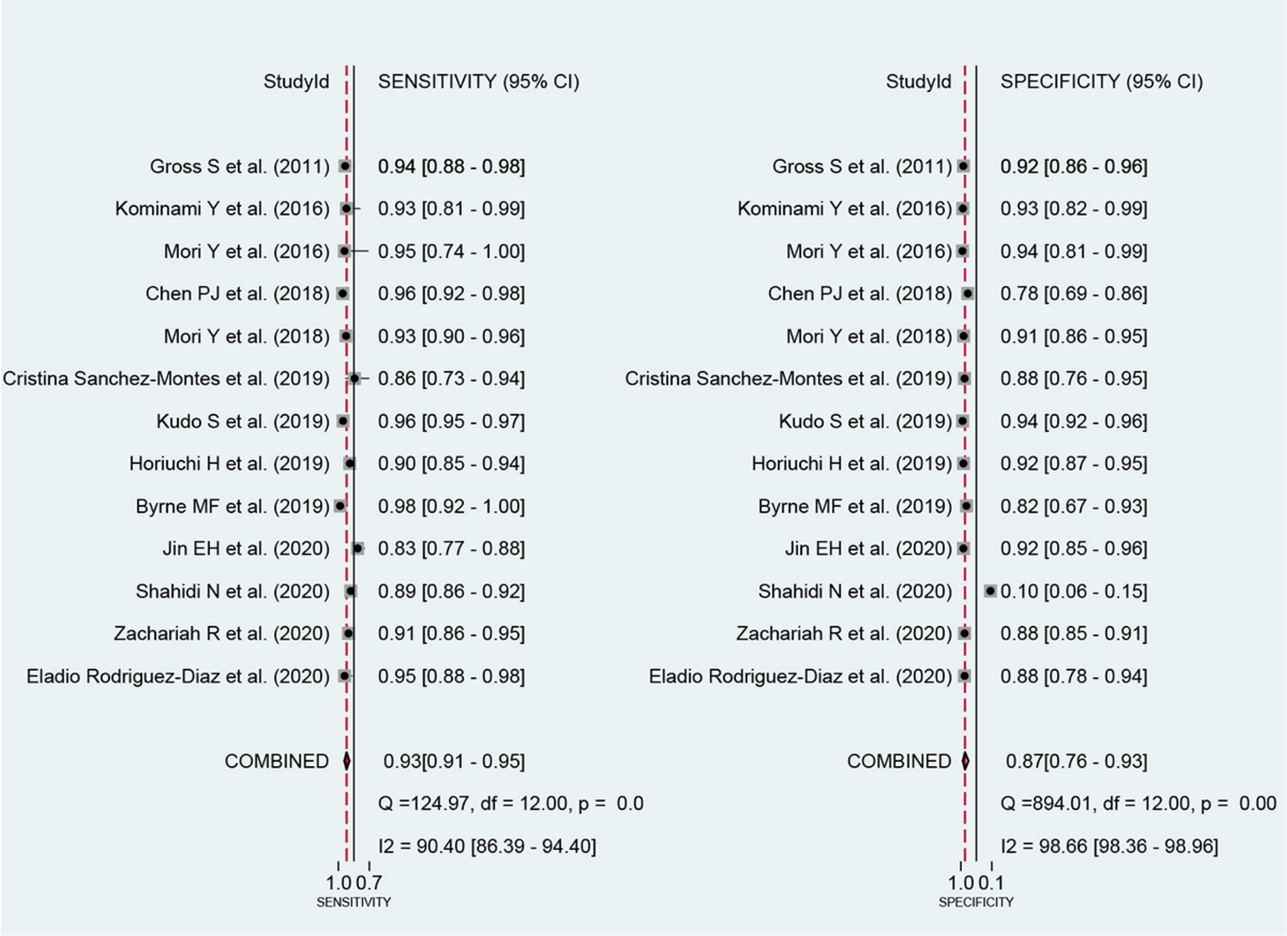


Table 2. Summary of diagnostic test accuracy meta-analysis and subgroup analysis for the diagnosis of diminutive colorectal polyps of the included studies.

Subgroup	Included studies (n=13), n (%)	AUC[a], mean (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PLR[b], mean (95% CI)	NLR[c], mean (95% CI)	DOR[d] (95% CI)
Value of meta-analysis in all the included studies	13 (100)	0.96 (0.93-0.97)	0.93 (0.91-0.95)	0.87 (0.76-0.93)	7.1 (3.8-13.3)	0.08 (0.06-0.11)	87 (38-201)
Comparative performance of CAD[e] models and endoscopists							
Value of CAD models in the comparative analysis	4 (31)	0.96 (0.94-0.98)	0.96 (0.95-0.97)	0.91 (0.84-0.95)	10.5 (5.7-19.1)	0.05 (0.03-0.06)	231 (113-473)
Value of expert endoscopists in the comparative analysis	4 (31)	0.97 (0.95-0.98)	0.93 (0.89-0.96)	0.89 (0.80-0.95)	8.8 (4.7-16.7)	0.08 (0.05-0.12)	116 (80-168)
Value of novice endoscopists in the comparative analysis	4 (31)	0.85 (0.82-0.88)	0.78 (0.68-0.86)	0.78 (0.68-0.86)	3.6 (2.3-5.8)	0.28 (0.18-0.43)	13 (6-30)
Methodological quality of included studies							
High	9 (69)	0.97 (0.95-0.98)	0.93 (0.90-0.95)	0.91 (0.88-0.93)	10.2 (7.7-13.5)	0.08 (0.05-0.11)	132 (83-211)
Low	4 (31)	0.92 (0.89-0.94)	0.93 (0.88-0.96)	0.87 (0.84-0.90)	7.4 (5.9-9.3)	0.08 (0.04-0.14)	96 (52-180)
Nationality of data							
Western	6 (46)	0.93 (0.90-0.95)	0.92 (0.90-0.94)	0.80 (0.51-0.94)	4.6 (1.6-13.7)	0.10 (0.06-0.16)	47 (11-213)
Asian	7 (54)	0.97 (0.95-0.98)	0.93 (0.89-0.96)	0.91 (0.87-0.94)	10.7 (7.4-15.3)	0.08 (0.05-0.12)	141 (80-248)
Type of test data sets							
Internal test	12 (92)	0.95 (0.93-0.96)	0.92 (0.89-0.94)	0.86 (0.75-0.93)	6.8 (3.5-13.3)	0.09 (0.06-0.13)	76 (32-179)
External test	2 (15)	Null	0.95 (0.94-0.96)	0.92 (0.90-0.93)	11.1 (4.7-26.4)	0.06 (0.03-0.15)	174 (36-841)
Location of polyps							
All	12 (92)	0.96 (0.93-0.97)	0.93 (0.90-0.95)	0.87 (0.75-0.93)	7.0 (3.6-13.9)	0.08 (0.06-0.11)	89 (36-220)
Rectosigmoid colon	3 (23)	0.97 (0.94-0.99)	0.91 (0.87-0.94)	0.91 (0.89-0.93)	6.8 (3.5-13.3)	0.09 (0.06-0.13)	76 (32-179)
Total number of included images							
≥200	8 (61)	0.95 (0.92-0.96)	0.92 (0.90-0.95)	0.84 (0.65-0.94)	5.9 (2.4-14.7)	0.09 (0.06-0.13)	66 (20-222)
<200	5 (38)	0.96 (0.94-0.98)	0.94 (0.90-0.96)	0.89 (0.84-0.93)	7.9 (5.5-11.2)	0.08 (0.04-0.15)	114 (57-230)
≥300	6 (46)	0.94 (0.91-0.95)	0.91 (0.88-0.94)	0.84 (0.57-0.96)	5.9 (1.8-19.6)	0.10 (0.06-0.17)	57 (12-275)
<300	7 (54)	0.97 (0.95-0.98)	0.94 (0.92-0.96)	0.88 (0.83-0.92)	8.1 (5.7-11.7)	0.06 (0.04-0.09)	127 (80-203)
Type of CAD models							
Neural network	6 (46)	0.94 (0.92-0.96)	0.93 (0.88-0.96)	0.76 (0.48-0.92)	4.0 (1.5-10.6)	0.09 (0.05-0.18)	42 (10-184)
SVM[f]	6 (46)	0.97 (0.96-0.98)	0.94 (0.91-0.96)	0.92 (0.90-0.94)	12.1 (9.0-16.5)	0.07 (0.04-0.10)	186 (101-344)
Type of endoscopic image							
Endocytoscope	3 (23)	0.98 (0.94-0.99)	0.95 (0.94-0.97)	0.94 (0.92-0.95)	13.8 (9.8-19.5)	0.05 (0.04-0.08)	248 (109-566)

Subgroup	Included studies (n=13), n (%)	AUC[a], mean (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PLR[b], mean (95% CI)	NLR[c], mean (95% CI)	DOR[d] (95% CI)
Endoscopy	10 (77)	0.95 (0.92-0.96)	0.92 (0.89-0.94)	0.85 (0.70-0.93)	6.0 (2.8-12.7)	0.09 (0.06-0.14)	64 (24-169)

[a] AUC: area under the curve.

[b] PLR: positive likelihood ratio.

[c] NLR: negative likelihood ratio.

[d] DOR: diagnostic odds ratio.

[e] CAD: computer-aided diagnosis.

[f] SVM: support vector machine.

Figure 5. SROC curve with a 95% confidence region and the prediction region of computer-aided diagnosis models for the diagnosis of histology for diminutive colorectal polyps in endoscopic images. AUC: area under the curve; SENS: sensitivity; SPEC: specificity; SROC: summary receiver operating characteristic.

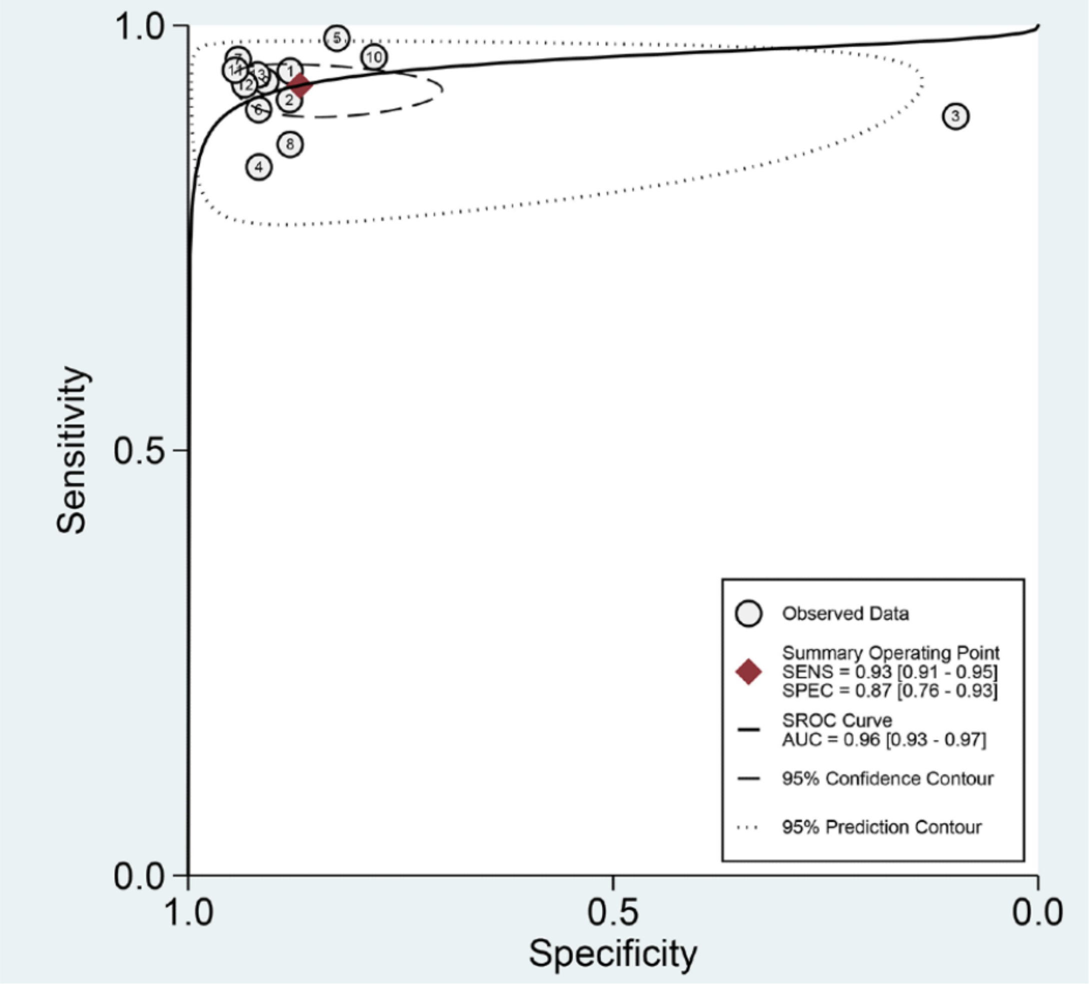
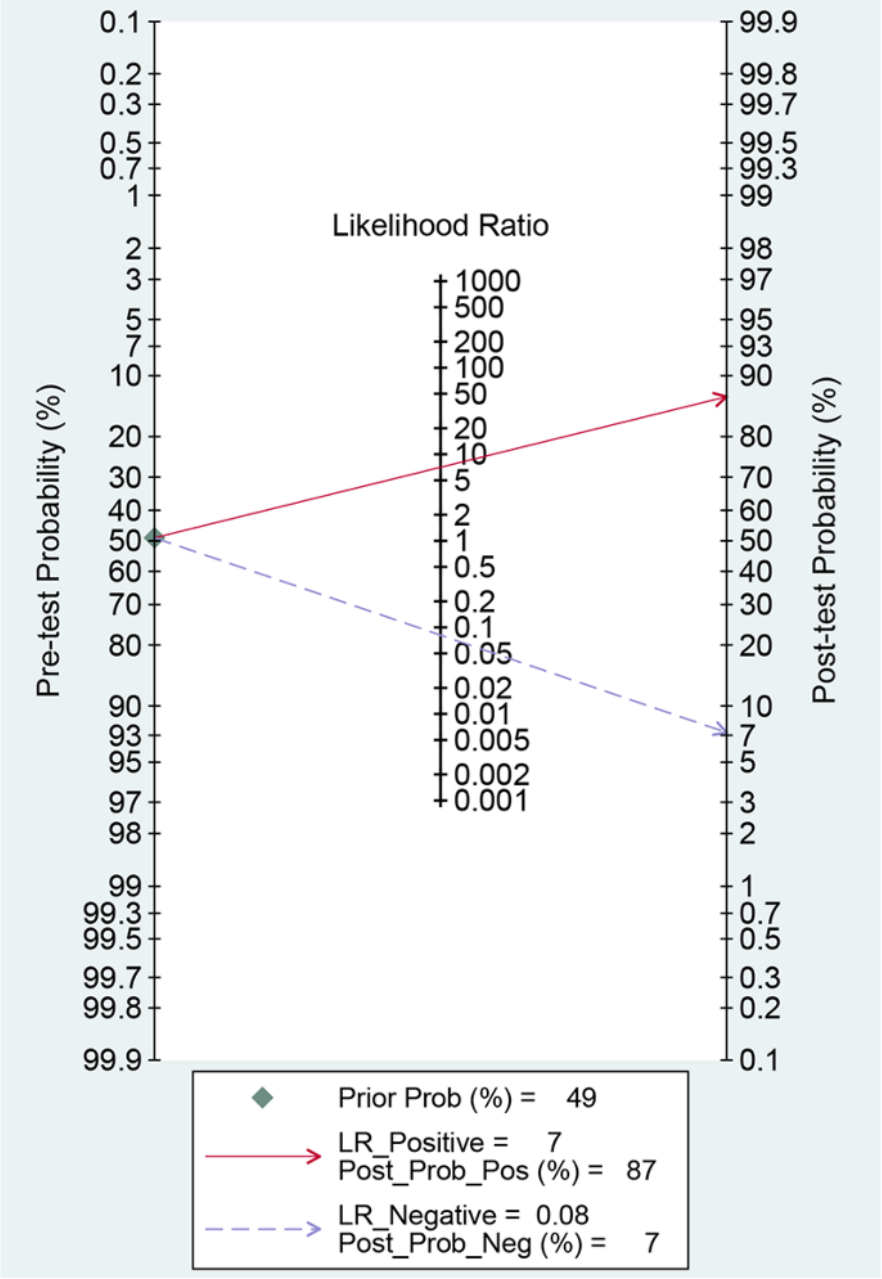


Figure 6. The Fagan nomogram for the diagnosis of histology for diminutive colorectal polyps in endoscopic images. LR: likelihood ratio; Post_Prob_Pos: the posterior probability of adenomas or neoplasms if the finding of the model was positive; Post_Prob_Neg: the posterior probability of adenomas or neoplasms if the finding of the model was negative.



Five studies [25,27,30,31,33] compared the performance of CAD models and endoscopists. Among these, four studies [27,30,31,33] have presented comparative performance between CAD models and endoscopists according to the expertise of the endoscopists (expert endoscopists vs CAD models or novice endoscopists vs CAD models). The pooled AUC, sensitivity, specificity, and DOR of CAD models for the diagnosis of DCPs were 0.96 (95% CI 0.94-0.98), 0.96 (95% CI 0.95-0.97), 0.91 (95% CI 0.84-0.95), and 231 (95% CI 113-473), respectively. For the expert endoscopists, the pooled AUC, sensitivity, specificity, and DOR were 0.97 (95% CI 0.95-0.98), 0.93 (95% CI 0.89-0.96), 0.89 (95% CI 0.80-0.95), and 116 (95% CI 80-168), respectively. For the novice endoscopists, the pooled AUC, sensitivity, specificity, and DOR were 0.85 (95% CI 0.82-0.88), 0.78 (95% CI 0.68-0.86), 0.78 (95% CI 0.68-0.86), and 13 (95% CI 6-30), respectively. The forest plot of AUCs is illustrated in Figure 7, and no significant difference was found between CAD models and expert endoscopists; however, novice endoscopists showed lower pooled AUC for the histologic diagnosis of DCPs than those for CAD models or expert endoscopists.

With regard to the NPV of CAD models for the diagnosis of adenomatous polyps in the rectosigmoid colon, the pretest prevalence of adenomatous polyp in the rectosigmoid colon was 13.2% (95% CI 10.2%-16.5%) in a recent meta-analysis [35]. For the assumption of this prevalence, the NPV of CAD models was 0.99 (95% CI 0.87-0.99; Figure 8). In this meta-analysis, the prevalence of adenomatous polyp in the rectosigmoid colon was 30% (95% CI 27%-32%) based on 29.53% (352/1192) of polyps in the rectosigmoid colon. For the assumption of this prevalence, the NPV of CAD models was 0.97 (95% CI 0.87-0.99; Figure 9). If we adopt a simple follow-up equation for the NPV of CAD models for the diagnosis of adenomatous polyps in the rectosigmoid colon using pooled sensitivity and specificity, the NPV of CAD models was 0.96 (95% CI 0.95-0.97). The follow-up equation is as follows:

NPV = (specificity × [1−prevalence])/(specificity × [1−prevalence] + prevalence × [1−sensitivity])

Figure 7. Forest plot of the area under the curve showing the comparative performance between computer-aided diagnosis models and endoscopists for the diagnosis of histology for diminutive colorectal polyps in endoscopic images. CAD: computer-aided diagnosis.

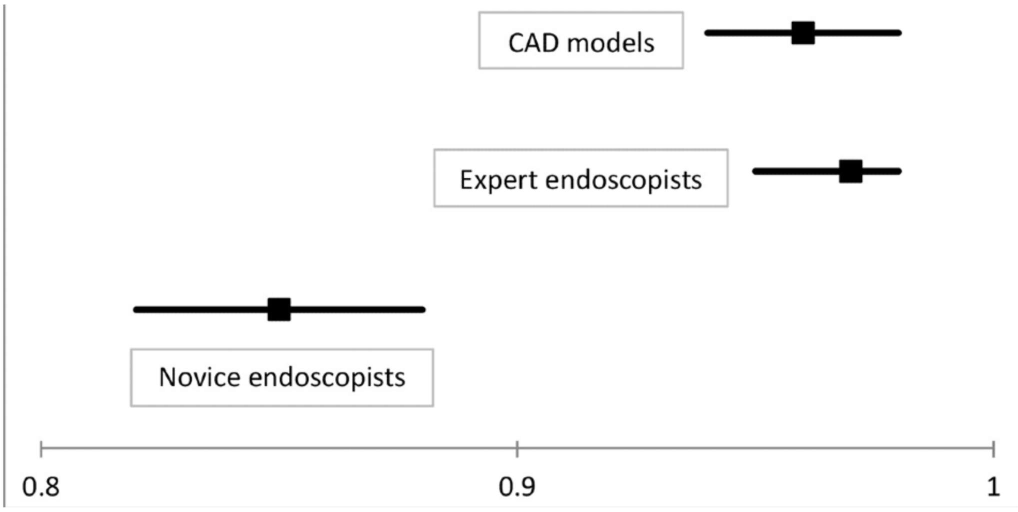


Figure 8. Probability-modifying plot of computer-aided diagnosis models for the diagnosis of adenomatous polyps in the rectosigmoid colon using endoscopic images (assumption of a prevalence of 13.2%). LR: likelihood ratio; NPV: negative predictive value; PPV: positive predictive value.

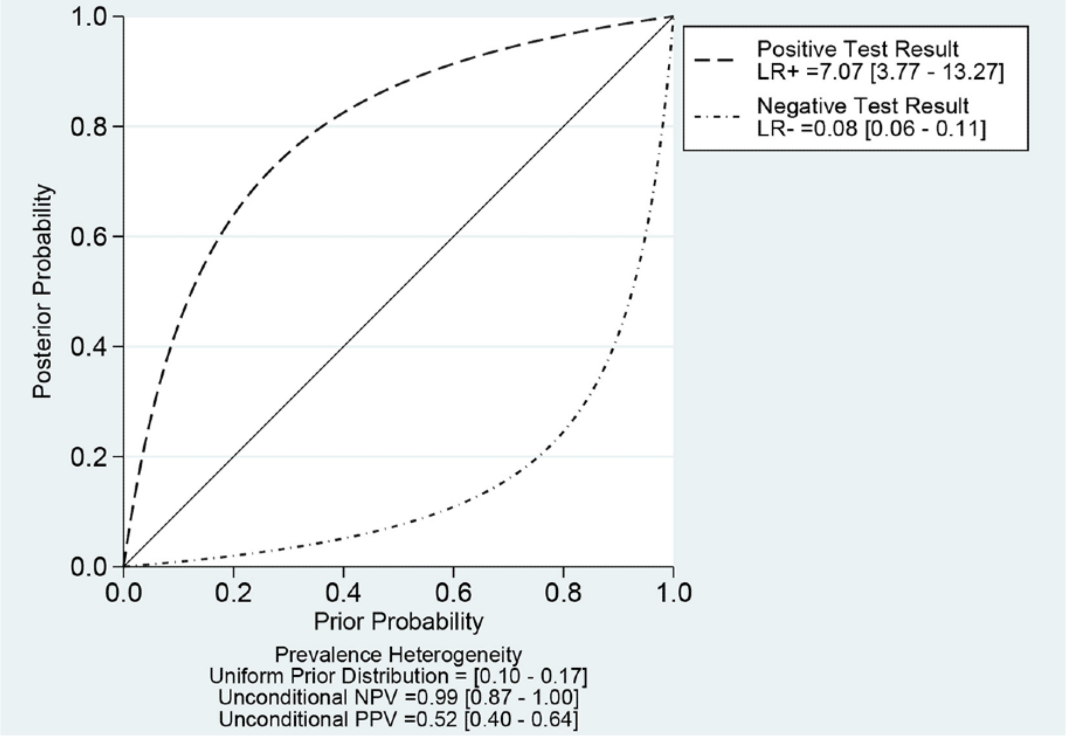
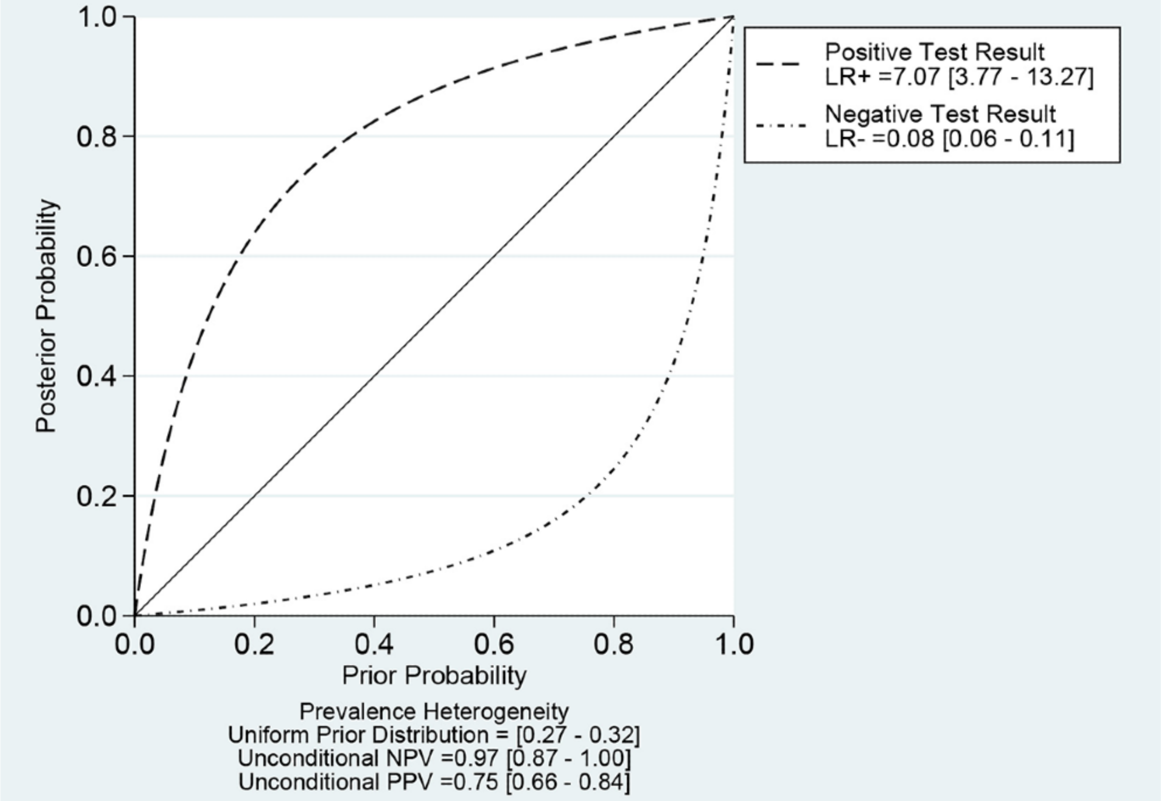


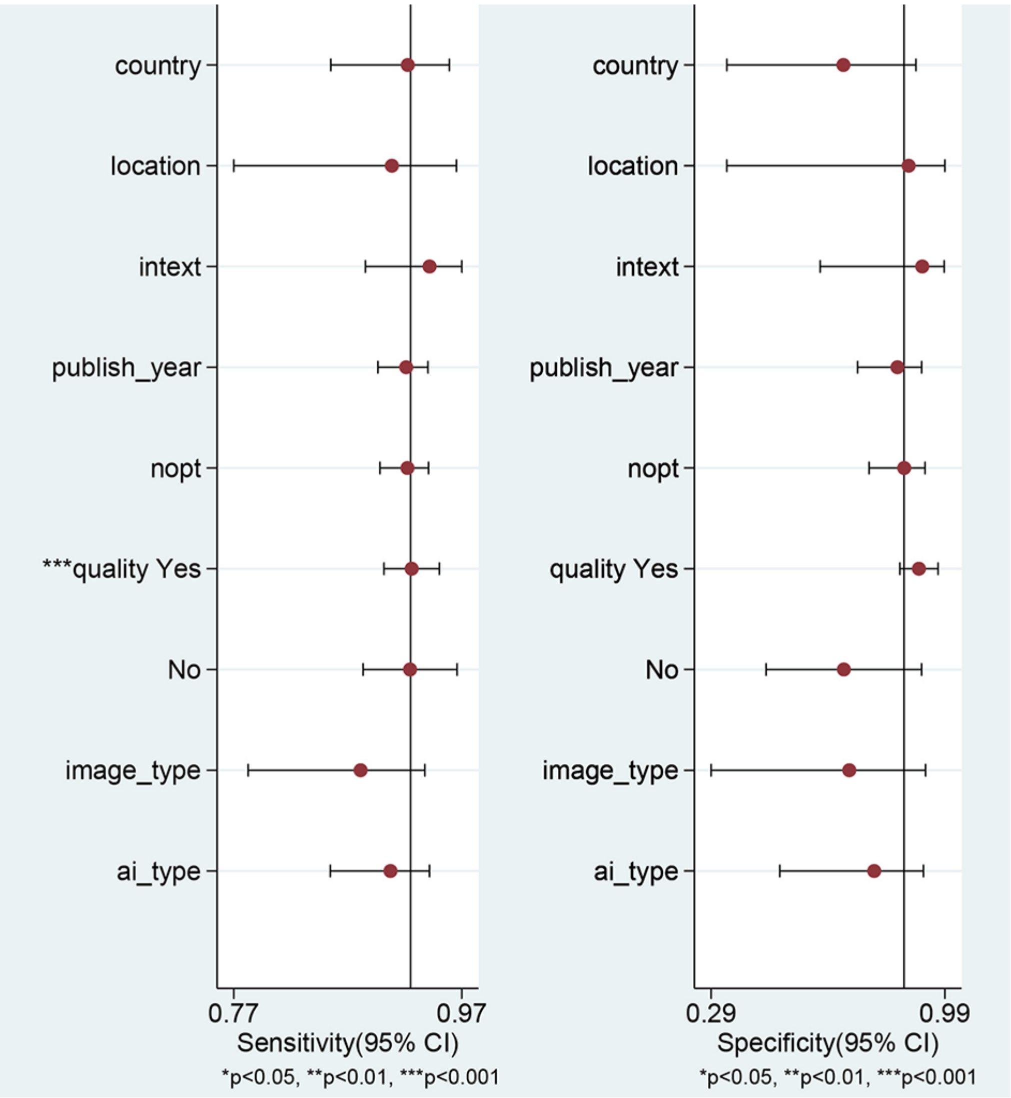
Figure 9. Probability-modifying plot of computer-aided diagnosis models for the diagnosis of adenomatous polyps in the rectosigmoid colon using endoscopic images (assumption of a prevalence of 30%). LR: likelihood ratio; NPV: negative predictive value; PPV: positive predictive value.



Assessment of Heterogeneity With Meta-Regression and Subgroup Analysis

First, the authors observed a positive correlation coefficient between the logit-transformed sensitivity and specificity (*r*=0.22) and an asymmetric β parameter, with a significant *P* value (*P*=.004), implying that heterogeneity exists among the studies. Second, a coupled forest plot of sensitivity and specificity was obtained (Figure 4). Compared with the enrolled studies, the study by Shahidi et al [23] showed lower specificity. This study was found to have an unclear risk of bias in methodology quality assessment. Therefore, subgroup analysis was carried out according to the methodological quality, and a negative correlation coefficient was found between logit-transformed sensitivity and specificity (*r*=−0.13) and an asymmetric β parameter, with a nonsignificant *P* value (*P*=.63) in high-quality studies, indicating an absence of heterogeneity among the studies. Third, the shape of the SROC curve for CAD of DCPs using endoscopic images was symmetric (Figure 5). Fourth, meta-regression using modifiers identified in the systematic review was conducted, and no source of heterogeneity could be identified (published year, *P*=.34; nationality of the data sets, *P*=.29; type of CAD models, *P*=.38; type of endoscopic image, *P*=.23; location of the DCPs, *P*=.90; type of test data sets, *P*=.66; total number of images, *P*=.66; and methodological quality, *P*=.10; Figure 10). Finally, a subgroup analysis based on the potential modifiers was performed, and the pooled AUC of studies with high methodological quality was higher than that of studies with lower methodological quality. Except for this variable (methodological quality), no significant changes in diagnostic performance were found according to the modifiers (Table 2).

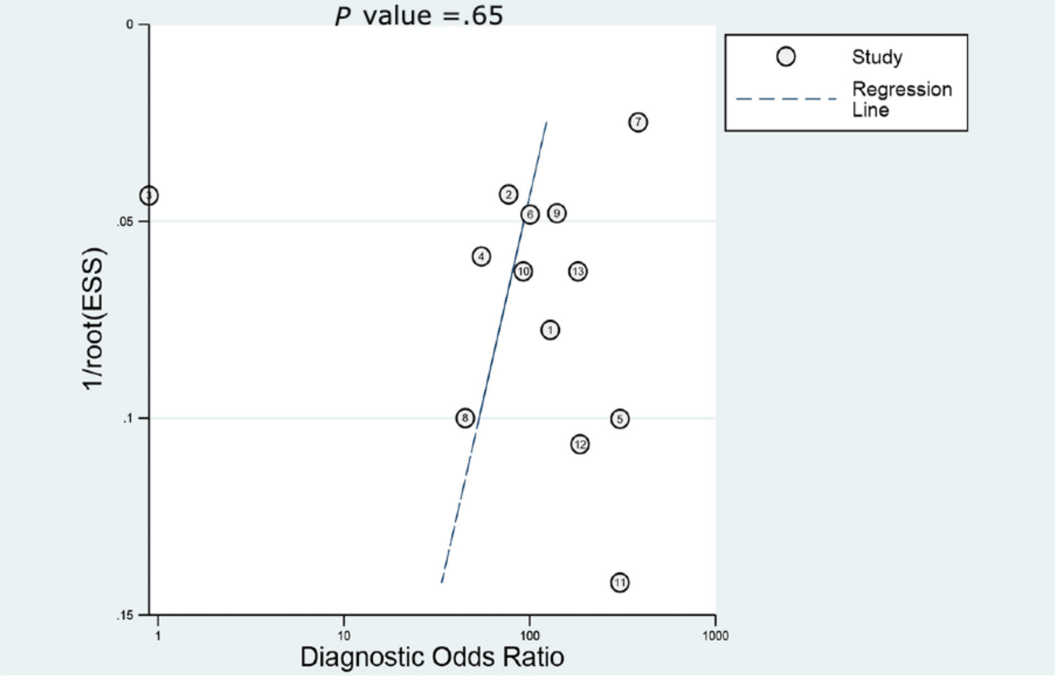
Figure 10. Univariable meta-regression plot of computer-aided diagnosis models for the diagnosis of histology for diminutive colorectal polyps using endoscopic images. ai: artificial intelligence; nopt: number of patients.



Evaluation of Publication Bias

Deek funnel plot of studies for the CAD of DCPs exhibited a symmetrical shape with respect to the regression line (Figure 11), and the asymmetry test showed no evidence of publication bias (*P*=.65).

Figure 11. Deek funnel plot of computer-aided diagnosis models for the diagnosis of histology for diminutive colorectal polyps using endoscopic images. ESS: explained sum of squares.



Discussion

Principal Findings

This study presented evidence that CAD models showed high performance values for the histologic diagnosis of DCPs and practical values in the Fagan nomogram, indicating the potential to use these models in clinical practice. This performance was comparable with that of expert endoscopists and higher than that of novice endoscopists. Although the main analysis found heterogeneity among the included studies, the subgroup analysis demonstrated that methodological quality was the reason for the heterogeneity. Thorough meta-regression or subgroup analyses did not reveal any additional reasons for heterogeneity.

Most polyps detected during colonoscopy are diminutive, and considering the low potential of malignancy, resecting all DCPs is not cost-effective [4,7-9]. However, many DCPs are still being resected and sent for histologic evaluation to determine the surveillance interval for CRC screening [10]. CAD models without pathologic diagnosis could lead to cost savings by changing surveillance interval recommendations, and the DTA meta-analysis in our study revealed that the NPV of CAD models for the diagnosis of adenomatous polyps in the rectosigmoid colon was over 90% to adopt the *diagnose and leave* strategy based on the optical biopsy satisfying PIVI performance thresholds for in situ endoscopic histology prediction of DCPs.

Despite the technical challenges of CAD models analyzing a smaller surface area, previous meta-analyses [36-40] have demonstrated that CAD models can increase the adenoma or polyp detection rate, especially for those with small size. With regard to the histologic prediction of DCPs, a previous meta-analysis revealed that endoscopists with only high confidence showed an NPV of approximately 90% using CAD models of digital chromoendoscopy, implicating the potential for adopting the *diagnosis and leave* strategy [41]. Another meta-analysis showed an NPV of 0.95 (95% CI 0.88-0.98) for the CAD of DCPs in nonmagnifying narrow-band imaging [42]. However, the location of the DCPs was not considered, and many studies were omitted from the search process.

An additional finding of this DTA meta-analysis is the robustness of the diagnostic performance of CAD models. The performance values were consistent regardless of the modifiers, except for the methodological quality (Table 2). This was consistent regardless of the nationality of the patients, location of DCPs, total number of included polyps, and type of CAD models or endoscopic images. However, the diagnostic performance of studies with high methodological quality showed higher AUCs than that of studies with low methodological quality, and no evidence of heterogeneity was detected among studies in the subgroup with high methodological quality. Although pooled AUCs in the subgroup of external test datasets could not be measured because only two studies were included in this subgroup, the remaining performance values were comparable with the subgroup of internal test data sets.

Limitations

Despite the robust evidence in the DTA meta-analysis stated earlier, several inevitable limitations were identified. First, only two or three studies were included in the subgroup analyses of external test data sets, rectosigmoid DCPs, and endocytoscopic images. Bivariate and HSROC methods are advanced statistical techniques that have overcome the limitations of the Moses-Shapiro-Littenberg method (which does not consider any heterogeneity between studies) [43,44]. However, the Moses-Shapiro-Littenberg method is only possible for a subgroup with fewer than four studies. With accumulating evidence on this topic, this statistical pitfall could be overcome. Second, the number of studies was insufficient to enable a comparison of the relative diagnostic performance of endoscopists and CAD models. Considering the real clinical adaptation of endoscopists with CAD models rather than endoscopists versus CAD models, it is no longer necessary to compare the diagnostic capabilities of doctors and CAD models [43]. Owing to the unique characteristics of patients in each institution, CAD models developed from a single institution usually have limitations for widespread implementation, indicating the importance of the external test. However, only two studies conducted external tests to verify CAD model performance. Additional studies focusing on external validation-oriented performance or suggesting a clinical application benefit for future perspectives in established CAD models are expected.

In conclusion, CAD models showed potential for the optical histological diagnosis of DCPs using endoscopic images.

Acknowledgments

This work was supported by the Technology Development Program (S2931703) and funded by the Ministry of SMEs and Startups (Korea).

Authors' Contributions

CSB was responsible for conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, supervision. CSB was also responsible for writing the original draft and reviewing and editing the final draft. JJL was responsible for data curation, formal analysis, investigation, and resources. GHB was responsible for data curation, formal analysis, investigation, and resources.

Conflicts of Interest

None declared.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021 Feb 04:209-249 [FREE Full text] [doi: 10.3322/caac.21660] [Medline: 33538338]
2. Zauber AG, Winawer SJ, O'Brien MJ, Lansdorp-Vogelaar I, van Ballegooijen M, Hankey BF, et al. Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. N Engl J Med 2012 Feb 23;366(8):687-696. [doi: 10.1056/nejmoa1100370]
3. Rex DK, Boland CR, Dominitz JA, Giardiello FM, Johnson DA, Kaltenbach T, et al. Colorectal cancer screening: recommendations for physicians and patients from the U.S. multi-society task force on colorectal cancer. Am J Gastroenterol 2017 Jul;112(7):1016-1030. [doi: 10.1038/ajg.2017.174] [Medline: 28555630]
4. Kandel P, Wallace MB. Should we resect and discard low risk diminutive colon polyps. Clin Endosc 2019 May;52(3):239-246 [FREE Full text] [doi: 10.5946/ce.2018.136] [Medline: 30661337]
5. Lieberman D, Moravec M, Holub J, Michaels L, Eisen G. Polyp size and advanced histology in patients undergoing colonoscopy screening: implications for CT colonography. Gastroenterology 2008 Oct;135(4):1100-1105 [FREE Full text] [doi: 10.1053/j.gastro.2008.06.083] [Medline: 18691580]
6. Qumseya BJ, Coe S, Wallace MB. The effect of polyp location and patient gender on the presence of dysplasia in colonic polyps. Clin Transl Gastroenterol 2012 Jul 26;3:e20 [FREE Full text] [doi: 10.1038/ctg.2012.14] [Medline: 23238292]
7. Butterly LF, Chase MP, Pohl H, Fiarman GS. Prevalence of clinically important histology in small adenomas. Clin Gastroenterol Hepatol 2006 Mar;4(3):343-348. [doi: 10.1016/j.cgh.2005.12.021] [Medline: 16527698]
8. Cho B, Bang CS. Artificial intelligence for the determination of a management strategy for diminutive colorectal polyps: hype, hope, or help. Am J Gastroenterol 2020 Jan;115(1):70-72. [doi: 10.14309/ajg.0000000000000476] [Medline: 31770118]
9. Gupta N, Bansal A, Rao D, Early DS, Jonnalagadda S, Wani SB, et al. Prevalence of advanced histological features in diminutive and small colon polyps. Gastrointest Endosc 2012 May;75(5):1022-1030. [doi: 10.1016/j.gie.2012.01.020] [Medline: 22405698]
10. Kaltenbach T, Anderson JC, Burke CA, Dominitz JA, Gupta S, Lieberman D, et al. Endoscopic removal of colorectal lesions-recommendations by the US multi-society task force on colorectal cancer. Gastroenterology 2020 Mar;158(4):1095-1129. [doi: 10.1053/j.gastro.2019.12.018] [Medline: 32122632]
11. Tao Pu L, Maicas G, Tian Y, Yamamura T, Nakamura M, Suzuki H, et al. Computer-aided diagnosis for characterization of colorectal lesions: comprehensive software that includes differentiation of serrated lesions. Gastrointest Endosc 2020 Oct;92(4):891-899. [doi: 10.1016/j.gie.2020.02.042] [Medline: 32145289]
12. Rex DK, Kahi C, O'Brien M, Levin T, Pohl H, Rastogi A, et al. The American Society for Gastrointestinal Endoscopy PIVI (Preservation and Incorporation of Valuable Endoscopic Innovations) on real-time endoscopic assessment of the histology of diminutive colorectal polyps. Gastrointest Endosc 2011 Mar;73(3):419-422. [doi: 10.1016/j.gie.2011.01.023] [Medline: 21353837]

13. Bang CS. [Deep learning in upper gastrointestinal disorders: status and future perspectives]. Korean J Gastroenterol 2020 Mar 25;75(3):120-131 [FREE Full text] [doi: 10.4166/kjg.2020.75.3.120] [Medline: 32209800]
14. Yang YJ, Bang CS. Application of artificial intelligence in gastroenterology. World J Gastroenterol 2019 Apr 14;25(14):1666-1683 [FREE Full text] [doi: 10.3748/wjg.v25.i14.1666] [Medline: 31011253]
15. McInnes MD, Moher D, Thombs BD, McGrath TA, Bossuyt PM, The PRISMA-DTA Group, et al. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: The PRISMA-DTA Statement. J Am Med Assoc 2018 Jan 23;319(4):388-396. [doi: 10.1001/jama.2017.19163] [Medline: 29362800]
16. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011 Oct 18;155(8):529-536. [doi: 10.7326/0003-4819-155-8-201110180-00009] [Medline: 22007046]
17. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol 2005 Oct;58(10):982-990. [doi: 10.1016/j.jclinepi.2005.02.022] [Medline: 16168343]
18. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. Stat Med 2001 Oct 15;20(19):2865-2884. [doi: 10.1002/sim.942] [Medline: 11568945]
19. Harbord RM, Whiting P. Metandi: Meta-analysis of diagnostic accuracy using hierarchical logistic regression. Stata J 2009 Aug 01;9(2):211-229. [doi: 10.1177/1536867x0900900203]
20. Littenberg B, Moses LE. Estimating diagnostic accuracy from multiple conflicting reports: a new meta-analytic method. Med Decis Making 1993;13(4):313-321. [doi: 10.1177/0272989X9301300408] [Medline: 8246704]
21. Rodriguez-Diaz E, Baffy G, Lo W, Mashimo H, Vidyarthi G, Mohapatra SS, et al. Real-time artificial intelligence-based histologic classification of colorectal polyps with augmented visualization. Gastrointest Endosc 2021 Mar;93(3):662-670. [doi: 10.1016/j.gie.2020.09.018] [Medline: 32949567]
22. Zachariah R, Samarasena J, Luba D, Duh E, Dao T, Requa J, et al. Prediction of polyp pathology using convolutional neural networks achieves "Resect and Discard" thresholds. Am J Gastroenterol 2020 Jan;115(1):138-144 [FREE Full text] [doi: 10.14309/ajg.0000000000000429] [Medline: 31651444]
23. Shahidi N, Rex DK, Kaltenbach T, Rastogi A, Ghalehjegh SH, Byrne MF. Use of endoscopic impression, artificial intelligence, and pathologist interpretation to resolve discrepancies between endoscopy and pathology analyses of diminutive colorectal polyps. Gastroenterology 2020 Feb;158(3):783-785. [doi: 10.1053/j.gastro.2019.10.024] [Medline: 31863741]
24. Jin EH, Lee D, Bae JH, Kang HY, Kwak M, Seo JY, et al. Improved accuracy in optical diagnosis of colorectal polyps using convolutional neural networks with visual explanations. Gastroenterology 2020 Jun;158(8):2169-2179. [doi: 10.1053/j.gastro.2020.02.036] [Medline: 32119927]
25. Byrne MF, Chapados N, Soudan F, Oertel C, Pérez ML, Kelly R, et al. Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model. Gut 2019 Jan 24;68(1):94-100 [FREE Full text] [doi: 10.1136/gutjnl-2017-314547] [Medline: 29066576]
26. Horiuchi H, Tamai N, Kamba S, Inomata H, Ohya TR, Sumiyama K. Real-time computer-aided diagnosis of diminutive rectosigmoid polyps using an auto-fluorescence imaging system and novel color intensity analysis software. Scand J Gastroenterol 2019 Jun;54(6):800-805. [doi: 10.1080/00365521.2019.1627407] [Medline: 31195905]
27. Kudo S, Misawa M, Mori Y, Hotta K, Ohtsuka K, Ikematsu H, et al. Artificial intelligence-assisted system improves endoscopic identification of colorectal neoplasms. Clin Gastroenterol Hepatol 2020 Jul;18(8):1874-1881. [doi: 10.1016/j.cgh.2019.09.009] [Medline: 31525512]
28. Sánchez-Montes C, Sánchez FJ, Bernal J, Córdova H, López-Cerón M, Cuatrecasas M, et al. Computer-aided prediction of polyp histology on white light colonoscopy using surface pattern analysis. Endoscopy 2019 Mar;51(3):261-265. [doi: 10.1055/a-0732-5250] [Medline: 30360010]
29. Mori Y, Kudo S, Misawa M, Saito Y, Ikematsu H, Hotta K, et al. Real-time use of artificial intelligence in identification of diminutive polyps during colonoscopy: a prospective study. Ann Intern Med 2018 Sep 18;169(6):357-366. [doi: 10.7326/M18-0249] [Medline: 30105375]
30. Chen P, Lin M, Lai M, Lin J, Lu HH, Tseng VS. Accurate classification of diminutive colorectal polyps using computer-aided analysis. Gastroenterology 2018 Feb;154(3):568-575. [doi: 10.1053/j.gastro.2017.10.010] [Medline: 29042219]
31. Mori Y, Kudo S, Chiu P, Singh R, Misawa M, Wakamura K, et al. Impact of an automated system for endocytoscopic diagnosis of small colorectal lesions: an international web-based study. Endoscopy 2016 Dec 5;48(12):1110-1118. [doi: 10.1055/s-0042-113609] [Medline: 27494455]
32. Kominami Y, Yoshida S, Tanaka S, Sanomura Y, Hirakawa T, Raytchev B, et al. Computer-aided diagnosis of colorectal polyp histology by using a real-time image recognition system and narrow-band imaging magnifying colonoscopy. Gastrointest Endosc 2016 Mar;83(3):643-649. [doi: 10.1016/j.gie.2015.08.004] [Medline: 26264431]
33. Gross S, Trautwein C, Behrens A, Winograd R, Palm S, Lutz HH, et al. Computer-based classification of small colorectal polyps by using narrow-band imaging with optical magnification. Gastrointest Endosc 2011 Dec;74(6):1354-1359. [doi: 10.1016/j.gie.2011.08.001] [Medline: 22000791]
34. Abadir AP, Ali MF, Karnes W, Samarasena JB. Artificial intelligence in gastrointestinal endoscopy. Clin Endosc 2020 Mar;53(2):132-141 [FREE Full text] [doi: 10.5946/ce.2020.038] [Medline: 32252506]

35. Wong MC, Huang J, Huang JL, Pang TW, Choi P, Wang J, et al. Global prevalence of colorectal neoplasia: a systematic review and meta-analysis. Clin Gastroenterol Hepatol 2020 Mar;18(3):553-561 [FREE Full text] [doi: 10.1016/j.cgh.2019.07.016] [Medline: 31323383]
36. Barua I, Vinsard DG, Jodal HC, Løberg M, Kalager M, Holme Ø, et al. Artificial intelligence for polyp detection during colonoscopy: a systematic review and meta-analysis. Endoscopy 2021 Mar 17;53(3):277-284. [doi: 10.1055/a-1201-7165] [Medline: 32557490]
37. Hassan C, Spadaccini M, Iannone A, Maselli R, Jovani M, Chandrasekar VT, et al. Performance of artificial intelligence in colonoscopy for adenoma and polyp detection: a systematic review and meta-analysis. Gastrointest Endosc 2021 Jan;93(1):77-85. [doi: 10.1016/j.gie.2020.06.059] [Medline: 32598963]
38. Zhang Y, Zhang X, Wu Q, Gu C, Wang Z. Artificial intelligence-aided colonoscopy for polyp detection: a systematic review and meta-analysis of randomized clinical trials. J Laparoendosc Adv Surg Tech A 2021 Feb 01:0777. [doi: 10.1089/lap.2020.0777] [Medline: 33524298]
39. Li J, Lu J, Yan J, Tan Y, Liu D. Artificial intelligence can increase the detection rate of colorectal polyps and adenomas: a systematic review and meta-analysis. Eur J Gastroenterol Hepatol 2021 Aug 01;33(8):1041-1048. [doi: 10.1097/MEG.0000000000001906] [Medline: 32804846]
40. Aziz M, Fatima R, Dong C, Lee-Smith W, Nawras A. The impact of deep convolutional neural network-based artificial intelligence on colonoscopy outcomes: a systematic review with meta-analysis. J Gastroenterol Hepatol 2020 Oct;35(10):1676-1683. [doi: 10.1111/jgh.15070] [Medline: 32267558]
41. Mason SE, Poynter L, Takats Z, Darzi A, Kinross JM. Optical technologies for endoscopic real-time histologic assessment of colorectal polyps: a meta-analysis. Am J Gastroenterol 2019 Aug;114(8):1219-1230. [doi: 10.14309/ajg.0000000000000156] [Medline: 30848728]
42. Lui TK, Guo C, Leung WK. Accuracy of artificial intelligence on histology prediction and detection of colorectal polyps: a systematic review and meta-analysis. Gastrointest Endosc 2020 Jul;92(1):11-22. [doi: 10.1016/j.gie.2020.02.033] [Medline: 32119938]
43. Bang CS, Lee JJ, Baik GH. Computer-aided diagnosis of esophageal cancer and neoplasms in endoscopic images: a systematic review and meta-analysis of diagnostic test accuracy. Gastrointest Endosc 2021 May;93(5):1006-1015 [FREE Full text] [doi: 10.1016/j.gie.2020.11.025] [Medline: 33290771]
44. Bang CS, Lee JJ, Baik GH. Artificial intelligence for the prediction of helicobacter pylori infection in endoscopic images: systematic review and meta-analysis of diagnostic test accuracy. J Med Internet Res 2020 Sep 16;22(9):e21983 [FREE Full text] [doi: 10.2196/21983] [Medline: 32936088]

Abbreviations

AUC: area under the curve
CAD: computer-aided diagnosis
CRC: colorectal cancer
DCP: diminutive colorectal polyp
DOR: diagnostic odds ratio
DTA: diagnostic test accuracy
FN: false-negative
FP: false-positive
HSROC: hierarchical summary receiver operating characteristic
NPV: negative predictive value
PIVI: Preservation and Incorporation of Valuable Endoscopic Innovations
PRISMA: Preferred Reporting Items for Systematic Review and Meta-Analyses
SROC: summary receiver operating characteristic
TN: true-negative
TP: true-positive

Edited by G Eysenbach; submitted 16.04.21; peer-reviewed by JA Benítez-Andrades, MT García-Ordás; comments to author 02.07.21; revised version received 04.07.21; accepted 27.07.21; published 25.08.21.

Please cite as:
Bang CS, Lee JJ, Baik GH
Computer-Aided Diagnosis of Diminutive Colorectal Polyps in Endoscopic Images: Systematic Review and Meta-analysis of Diagnostic Test Accuracy
J Med Internet Res 2021;23(8):e29682
URL: https://www.jmir.org/2021/8/e29682
doi:10.2196/29682
PMID:34432643

©Chang Seok Bang, Jae Jun Lee, Gwang Ho Baik. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 25.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Patient Perceptions on Data Sharing and Applying Artificial Intelligence to Health Care Data: Cross-sectional Survey

Ravi Aggarwal[1], MBBS; Soma Farag[1], BSc; Guy Martin[1], BSc, MBBS, PhD; Hutan Ashrafian[1], MBBS, PhD; Ara Darzi[1], MD

Institute of Global Health Innovation, Imperial College London, London, United Kingdom

Corresponding Author:
Hutan Ashrafian, MBBS, PhD
Institute of Global Health Innovation
Imperial College London
10th Floor, QEQM Building,
St Marys Hospital, Praed St
London, W2 1NY
United Kingdom
Phone: 44 07799871597
Email: h.ashrafian@imperial.ac.uk

Abstract

Background: Considerable research is being conducted as to how artificial intelligence (AI) can be effectively applied to health care. However, for the successful implementation of AI, large amounts of health data are required for training and testing algorithms. As such, there is a need to understand the perspectives and viewpoints of patients regarding the use of their health data in AI research.

Objective: We surveyed a large sample of patients for identifying current awareness regarding health data research, and for obtaining their opinions and views on data sharing for AI research purposes, and on the use of AI technology on health care data.

Methods: A cross-sectional survey with patients was conducted at a large multisite teaching hospital in the United Kingdom. Data were collected on patient and public views about sharing health data for research and the use of AI on health data.

Results: A total of 408 participants completed the survey. The respondents had generally low levels of prior knowledge about AI. Most were comfortable with sharing health data with the National Health Service (NHS) (318/408, 77.9%) or universities (268/408, 65.7%), but far fewer with commercial organizations such as technology companies (108/408, 26.4%). The majority endorsed AI research on health care data (357/408, 87.4%) and health care imaging (353/408, 86.4%) in a university setting, provided that concerns about privacy, reidentification of anonymized health care data, and consent processes were addressed.

Conclusions: There were significant variations in the patient perceptions, levels of support, and understanding of health data research and AI. Greater public engagement levels and debates are necessary to ensure the acceptability of AI research and its successful integration into clinical practice in future.

(J Med Internet Res 2021;23(8):e26162) doi:10.2196/26162

KEYWORDS

artificial intelligence; patient perception; data sharing; health data; privacy

Introduction

Recent advances in data science and artificial intelligence (AI) technologies have the potential to transform the way patient-centered health care is delivered [1]. AI is a branch of computer science that refers to the ability of computers or machines to creatively solve problems that would normally require human intelligence. Machine learning (ML) is a subset of AI that provides systems with the ability to automatically learn and improve from experience without explicitly being programmed. It involves algorithms that are designed to emulate human intelligence by learning from their environment [2]. Considerable research is being conducted as to how AI and ML can be applied to health care, with diagnostics seeming to be the most promising field for AI implementation [3,4]. However, for AI research to be successful and truly translational, large amounts of health data are required for training and testing algorithms [5]. Therefore, public trust and support for using health data in AI research are essential.

Public perceptions regarding sharing of health data for research are well characterized [6-8]. Although concerns regarding the privacy, confidentiality, and commercial motives associated with data sharing are frequently highlighted, when people perceive that public or societal benefits arise from such research and when they place trust in the organizations conducting the research, they are generally supportive [7]. However, patient and public perceptions regarding health data sharing for AI research are not sufficiently characterized [9]. Data sharing for AI research purposes is a controversial subject, and therefore, conditional public support for data sharing cannot be assumed to extend to this field of research [10]. Reasons for this include knowledge and understanding of AI in general [10], ethical concerns [11], and fears around the potential reidentification of anonymized personal health data [12]. Furthermore, recent negative media reports about large technology companies using health data for AI research [13] and several important data breaches and cyberattacks [14] may undermine public trust in this technology.

Despite these additional issues, there is limited research exploring patient perceptions on data sharing for AI research purposes [10,15-18]. If the promises of AI are to be truly realized in health care, strategic public debates are important to ensure that the public maintains trust in the technology and use of confidential health data [19]. This is now especially important as regulatory approval has already been granted for AI-powered diagnostic software to be used in routine clinical practice [20].

Therefore, the aim of this study was to survey a large sample of patients at our hospital to identify their current awareness on health data research, and viewpoints on data sharing for AI research purposes and using AI technology on health care data.

Methods

Survey Development

We conducted a cross-sectional study using a self-completed questionnaire survey tool with patients at a large, multisite university teaching hospital in London. The survey tool was developed via a multistep codesign process in collaboration with patients. First, a literature review was conducted to identify the initial survey themes and items, which were then used to inform the codesigning process of a prototype questionnaire with a patient focus workshop. The workshop was a 3-hour face-to-face meeting with subject matter experts and a group of 3 patients selectively chosen out of 9 individuals who applied. The patients were chosen for their experience in survey development and had previously been involved in research studies at our organization. The feedback and suggestions from the workshop were analyzed by two researchers (RA and HA) and changes were made to the prototype questionnaire based on this feedback. The revised survey was then emailed to the workshop participants for further review with no more changes suggested. Finally, a pilot study was conducted with 5 patients of varying ages, genders, education levels, and ethnicities recruited opportunistically from an outpatient clinic in our hospital to evaluate comprehension and measure the average time taken to complete the survey. We were able to ascertain that all patients understood the information sheet and the questions, and they were able to complete the survey within 12 minutes.

Sample

The participants were opportunistically recruited from outpatient waiting areas or from the inpatient wards over a 12-week period beginning June 2018. The eligibility criteria for participation were as follows: (1) 16 years or older, (2) able to understand the information describing the research study, and (3) willing and able to provide informed written consent. The study was reviewed and approved by the South East Scotland Research Ethics Service (18/SS/0057/AM01).

Data were collected on patient and public views about sharing health data for research and the use of AI on health data. The front page of the questionnaire introduced the participants to AI, electronic health records, and data anonymization and sharing. The participants were informed about the aims of the questionnaire, and they voluntarily participated after being given a patient information sheet and the opportunity to ask questions. Patient anonymity was ensured, and the responses were identified by participant identification numbers only. The 24-item questionnaire examined various aspects related to patient and public views on the subject and was split into 4 sections:

1. awareness of health data usage for research
2. views on data sharing, consent, and anonymization
3. views on AI
4. sociodemographic characteristics and health statuses of the participants

Statistical Analysis

All the surveys were completed on paper before being manually entered into a database in Microsoft Excel (Microsoft Corporation). Descriptive statistics were used to describe the sample by gender, age, ethnicity, educational attainment, perceived health status, Internet usage, and smartphone ownership. The age categories included 16-30, 31-45, 46-64, and 65+. Educational attainments were classified as “low” (General Certificate of Secondary Education [GCSE] or below), “medium” (Advanced Certificate of Secondary Education [A-Level] or equivalent) or “high” (university degree and above). Ethnicities were grouped as either “Caucasian” (White/British or White/Other) or “Black, Asian, and minority ethnic (BAME)” (African/Caribbean, Asian, mixed or multiple ethnicities, or other). Personal health statuses were classified as “high” (good, very good, or excellent) or “low” (poor or fair). Internet usage was categorized as “daily” or “less frequent/no access” and smartphone ownership as “yes” or “no/prefer not to say.”

For questions with Likert-type ordinal responses, ordinal logistic regression was performed to examine the relationships between the responses and the demographic variables mentioned above. Binary logistic regression was used for questions with binary responses. These methods were used because of the nature of the dependent and independent variables and because they could be adjusted for other demographic variables, and any confounding effects could be removed. For each demographic variable, the categories were compared with a predefined reference group for performing logistic regression. The reference groups were "female" for the sex variable, 65+ for age, BAME for ethnicity, "high" for education level, "low" for personal health status, "less frequent/no access" for Internet usage, and "no" for smartphone ownership. The results were deemed statistically significant if *P*<.05. Statistical analysis was performed using SPSS (version 27.0, IBM Corp).

Data Sharing

Access to deidentified data might be provided on reasonable request when accompanied by a study protocol and analysis plan. Requests are subject to the establishment of appropriate data governance and approval by a committee involving the current research team. Requests must be made in writing to the corresponding author.

Results

Participants

A total of 408 participants recruited from all 5 sites of a multicenter university teaching hospital in the United Kingdom completed the survey. The demographic characteristics of the respondents are presented in Table 1. Internet usage (59/60, 98.9% in the 16-30 group mentioning daily usage compared to 48/61, 78.9% in the 65+ group) and smartphone ownership (59/60, 98.9% in the 16-30 group compared to 35/61, 57.9% in the 65+ group) declined with increasing age. Daily Internet usage reduced with reducing educational attainments (158/167, 94.6% in the "high" group [university degree and above] compared to 56/67, 83.1% in the "low" group [GCSE and below]). Similarly, smartphone use decreased with decreasing educational attainment (158/167, 94.6% in the "high" group compared with 48/67, 71.5% in the "low" group). Moreover, 90.9% (286/315) of smartphone users used the Internet daily compared to 75.5% (40/53) of non-smartphone Internet users.

The full breakdown of the questions and answers are given in Tables 2 and 3. Figure 1 shows a significance map with details on the directionality and level of significance associated with the responses and all the demographic variables (see Multimedia Appendix 1 for the results of the logistic regression analyses).

Table 1. Demographic characteristics of the respondents.

Characteristic	Number of respondents (N=408)	Percentage (%)
Gender		
Male	173	42.4
Female	198	48.5
Unanswered	37	9.1
Age (years)		
16-30	90	22.1
31-45	81	19.9
46-64	123	30.1
65-79	61	15
>80	15	3.7
Unanswered	38	9.3
Ethnicity		
White/British	174	42.6
White/Other	55	13.5
African/Caribbean	45	11
Asian	56	13.7
Mixed or multiple ethnic	10	2.5
Other	26	6.4
Unanswered	42	10.3
Education		
No qualifications	34	8.3
GCSE[a]/O-Level[b]/NVQ[c]	67	16.4
A-Level[d]	70	17.2
University degree	167	40.9
Other	29	7.1
Unanswered	41	10.0
Personal health status		
Poor	42	10.3
Fair	107	26.2
Good	119	29.2
Very good	86	21.1
Excellent	18	4.4
Unanswered	36	8.8
Internet usage		
Daily	301	73.8
Less frequent	42	10.3
No access	29	7.1
Unanswered	36	8.8
Smartphone ownership		
Yes	315	77.2
No	53	13.0

Characteristic	Number of respondents (N=408)	Percentage (%)
Prefer not to say	4	1.0
Unanswered	36	8.8

[a] GCSE: General Certificate of Secondary Education.

[b] O-Level: General Certificate of Education Ordinary Level

[b] NVQ: National Vocational Qualifications.

[c] A-Level: General Certificate of Education Advanced Level.

Table 2. Respondents' opinions on health data usage and sharing for research.

Question	Number of respondents	Responses				
1. How much would you say you know about how the following organizations use health data for research purposes?		Never heard of (%)	Heard of, know nothing about (%)	Just a little (%)	A fair amount (%)	A great deal (%)
a. NHS[a]	407	14.3	13.3	30.7	26.8	15
b. Commercial organizations	405	25.7	22.7	27.9	15.1	8.6
c. University researchers	405	18.5	21.7	32.1	19.5	8.1
2. How likely would you be to allow your anonymized health information to be used for the purposes of medical research by the following organizations?		Very unlikely (%)	Fairly unlikely (%)	Not sure (%)	Fairly likely (%)	Very likely (%)
a. NHS	408	4.4	3.9	13.7	30.4	47.5
b. Commercial organizations	405	22.0	17.5	34.1	14.1	12.3
c. University researchers	405	6.7	6.4	21.2	31.4	34.3
		Very unlikely (%)	Fairly unlikely (%)	Not sure (%)	Fairly likely (%)	Very likely (%)
3. To what extent would you support ICL[b] creating a large, anonymized set of data of routinely collected ICHNT[c] health care data for AI[d] research purposes?	408	1.5	2.5	12.3	48.5	35.3
		Very unlikely (%)	Fairly unlikely (%)	Not sure (%)	Fairly likely (%)	Very likely (%)
4. To what extent would you support the transfer of your anonymized health data to ICL if there was a very small chance of it being reidentified after transfer?	407	13.0	13.5	17.2	36.3	19.9
		Strongly disagree (%)	Disagree (%)	Neither agree nor disagreed (%)	Agree (%)	Strongly agree (%)
5. Currently, researchers are legally allowed to access anonymized health data for research without the need for patient consent. To what extent do you agree with this?	407	9.3	14.7	20.1	39.1	16.7

[a] NHS: National Health Service.

[b] ICL: Imperial College London

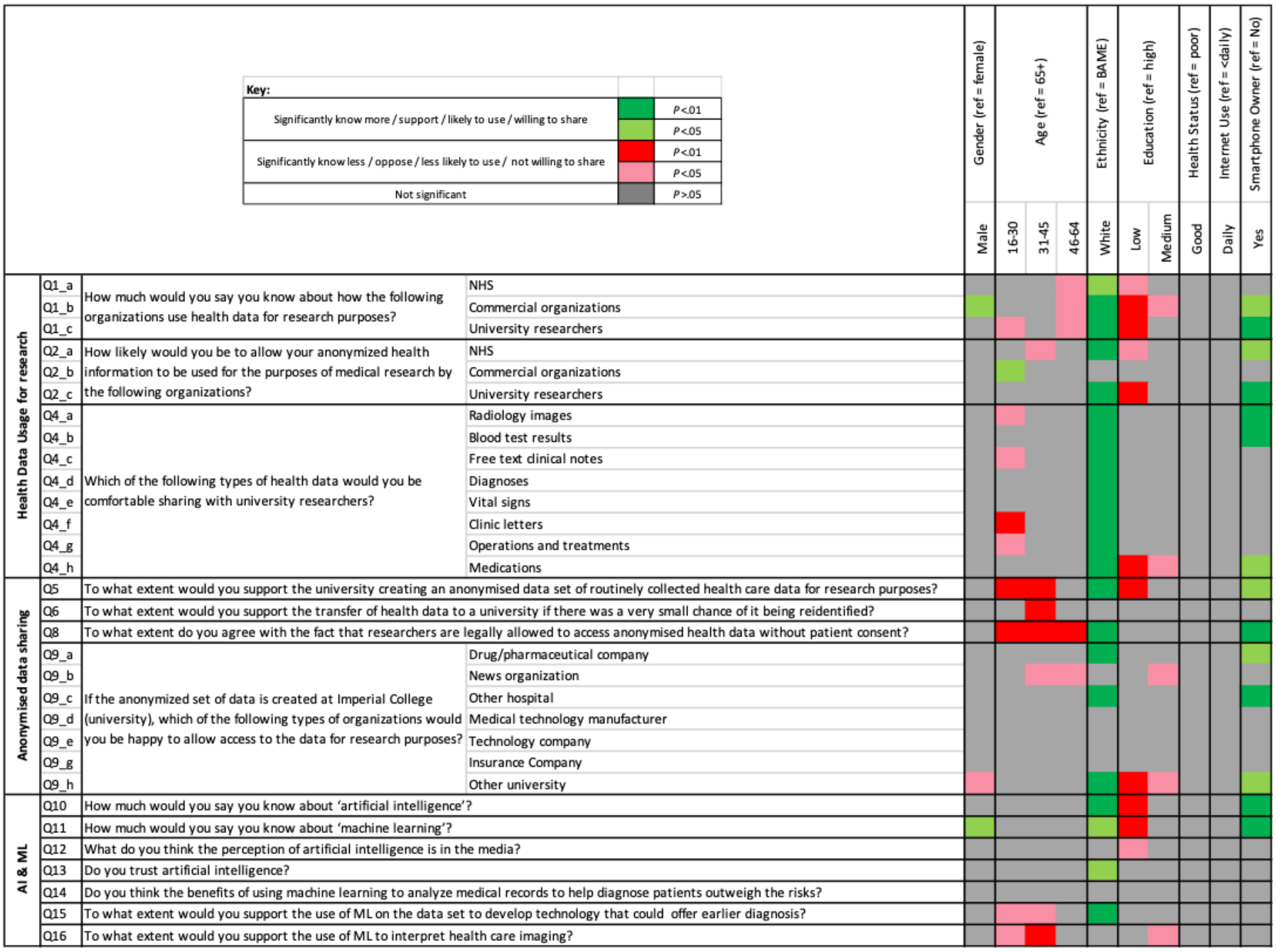
[c] ICHNT: Imperial College Healthcare NHS Trust.

[d] AI: artificial intelligence.

Table 3. Respondents' opinions on artificial intelligence and machine learning.

Question	Number of respondents	Responses				
		Never heard of (%)	Slightly aware (%)	Somewhat aware (%)	Moderately aware (%)	Extremely aware (%)
How much would you say you know about "artificial intelligence?"	407	22.9	20.4	20.9	29.2	6.6
		Never heard of (%)	Slightly aware (%)	Somewhat aware (%)	Moderately aware (%)	Extremely aware (%)
How much would you say you know about "machine learning?"	407	27.3	21.9	22.4	22.9	5.7
		Very negative (%)	Slightly negative (%)	Not sure (%)	Slightly positive (%)	Very positive (%)
What do you think the perception of artificial intelligence is in the media?	207	1.4	29.5	25.6	34.3	9.2
		Strongly distrust (%)	Slightly distrust (%)	Not sure (%)	Slightly trust (%)	Strongly trust (%)
Do you trust artificial intelligence?	206	3.4	12.6	29.5	43.5	10.6
		Risk outweighs benefits (%)	Risk and benefits equal (%)	Benefits outweighs risk (%)	Don't know (%)	
Do you think the benefits of using machine learning to analyze medical records to help diagnose patients outweighs the risks?	205	6.8	22.7	45.9	23.7	
		Strongly oppose (%)	Tend to oppose (%)	Neither support nor oppose (%)	Tend to support (%)	Strongly support (%)
To what extent would you support the use of machine learning to develop technology that could potentially offer earlier diagnosis and more accurate treatments to patients?	206	0.5	3.4	8.7	52.7	34.3
		Strongly oppose (%)	Tend to oppose (%)	Neither support nor oppose (%)	Tend to support (%)	Strongly support (%)
To what extent would you support the use of machine learning to interpret health care imaging as an aid for doctors when reporting these images?	206	1.0	2.9	9.7	48.1	38.3

Figure 1. Significance map detailing the directionality and significance of the relationships between the responses and the panel of demographic characteristics. BAME: Black, Asian, and minority ethnic; ML: machine learning; NHS: National Health Service.



Awareness of Health Data Usage for Research

NHS

Among the 407 respondents, 170 (41.7%) knew "a fair amount'" or "a great deal" about how the NHS uses health data for research purposes (Question 1a), and 318/408 (77.9%) were "fairly likely" or "very likely" to allow their anonymized health information to be used for medical research purposes by the NHS (Question 2a). In comparison with their reference group, those aged 31-45 (*P*=.013) and with lower educational attainment (*P*=.019) were significantly less likely to be comfortable sharing health data, whereas Caucasian groups (*P*<.001) and those who own smartphones (*P*=.014) were more likely to be comfortable sharing data with the NHS for research purposes.

Commercial Organizations

Only 96/405 (23.7%) knew "a fair amount" or "a great deal" about how commercial organizations use health data for research purposes (Question 1b), and 107/405 (26.4%) were "fairly likely" or "very likely" to allow their anonymized health information to be used for medical research purposes by commercial organizations (Question 2b). In comparison with their reference group, those aged 16-30 (*P*=.042) were significantly more likely to be comfortable sharing data with commercial organizations for research purposes.

University Researchers

Of the 405 respondents, 112 (27.7%) knew "a fair amount" or "a great deal" about how university researchers use health data for research purposes (Question 1c), and 266/405 (65.7%) were "fairly likely" or "very likely" to allow their anonymized health information to be used for medical research purposes by university researchers (Question 2c). In comparison with their reference group, those of lower educational attainment (*P*=.003) were significantly less likely to be comfortable sharing health data, whereas Caucasian groups (*P*<.001) and those owning smartphones (*P*=.007) were more likely to share data with university researchers.

As for the types of data shared with university researchers, over 70% of respondents were comfortable sharing information on radiology, blood test results, diagnoses, operations and treatments and medications (Question 4). However, fewer respondents were comfortable sharing clinic letters (51%), free text clinical notes (51.2%), or vital signs (67.2%). Caucasian respondents were significantly more likely to be comfortable sharing all data types (*P*=.001). Those under 30 were less likely to be comfortable sharing data on operations and treatments, free text clinical notes, and radiology images (all *P*<.05), and clinic letters (all *P*<.01). Smartphone owners were more likely to be comfortable sharing radiology images, blood test results (all *P*<.01), and medication data (all *P*<.05).

Data Sharing, Consent, and Anonymization

Among the 408 respondents, 342 (83.8%) “tend to support” or “strongly support” the creation of an anonymized data set of routinely collected NHS data for AI research purposes at the university partner (Question 5). In comparison to their reference counterpart, respondents under the age of 45 (*P*=.002) or having lower educational achievement (*P*=.003) were statistically less likely to support data set creation, whereas those of Caucasian background (*P*=.006) and smartphone owners (*P*=.033) were more likely to support this. Fewer respondents would support the transfer of anonymized routinely collected health data to a university partner if there was a small chance of reidentification after transfer (229/407, 56.2%) (Question 6). Those aged 31-45 were significantly less likely to support this when compared with the reference group (*P*=.008).

Furthermore, greater than 50% (227/407, 55.7%) of the respondents cited that individual-level patient consent should not be required to use anonymized routinely collected health care data for research purposes, as is the status quo (Question 8). All age groups below 65 were significantly less likely to agree with this compared with those over 65 (all *P*<.01). Those of Caucasian background (*P*<.001) and smartphone owners (*P*=.008) were more likely to agree.

With respect to allowing third party organizations access to anonymized data for research purposes, respondents were uncomfortable sharing data with news organizations (6.9%), insurance companies (6.9%), and technology companies (21.6%) (Question 9). Those aged 31-64 and with medium educational attainment were significantly less inclined to provide access to news organizations (all *P*<.05). Respondents were slightly more inclined to provide data access to drug/pharmaceutical companies (47.1%), medical technology companies (46.1%), other universities (44.1%), and other hospitals (68.9%). Caucasians were significantly more comfortable with providing access to these organizations. Females and those of low and medium educational attainments were significantly less likely to be comfortable sharing data with other universities.

AI and ML Research

More respondents were familiar with AI (231/407, 56.7%) than ML (207/407, 50.8%). Further, 22.9% (93/407) and 27.3% (111/407) had never heard of AI and ML, respectively. Patients from Caucasian backgrounds (*P*=.003, *P*=.028), males (*P*=.066, *P*=.025), and smartphone owners (*P*<.001, *P*<.001) were significantly more aware about AI and ML in comparison with their reference groups (Question 10). Those of lower educational attainment are significantly less familiar with these terminologies (*P*<.001, *P*<.001).

As we identified that 49.2% (200/407) of respondents stated they had “not heard of” or were only “slightly aware” of ML, the responses from those respondents were excluded from the results of questions 12-16. Moreover, 90/207 (43.5%) think that the perception of AI in the media is very positive or slightly positive and 112/206 (54.1%) of respondents strongly trust or slightly trust AI. Caucasians have significantly more trust in AI (*P*=.035) than BAME patients. Furthermore, 95/205 (45.9%) think that the benefits of AI in health care outweighed the risks compared with 6.8% (14/205) who think that the risks outweigh the benefits (Question 14). With regard to supporting ML research, 87.4% (180/206) and 86.4% (178/206) strongly support or tend to support this on anonymized health care data and health care imaging respectively (Questions 15 and 16). Caucasians were significantly more likely to support this research (*P*=.01), whereas those aged 16-30 and 31-45 were significantly less supportive of this research on health care data (*P*=.013 and *P*=.027 respectively).

Discussion

Major Findings

The increasing availability of health care data and exponential rise of computational power have caused the recent surge in AI applications in health care [5]. Powerful AI techniques can potentially assist physicians to make better clinical decisions or even perform some tasks autonomously. The successful integration and translation of this technology into routine clinical practice, depends not only on numerous technological challenges, but also whether the public and patients can accept and trust it [21].

In this study, which to the best of our knowledge is the first one assessing patients views about sharing health care data for AI research from a UK hospital, several key findings emerged. Consistent with previous literature [10], we found that patients report generally low levels of knowledge about AI and ML. This is a key finding; if the use of AI in healthcare is to increase, educating patients about the risks and benefits of this technology is crucial [19]. The vision of AI presented in the press and other forms of media [22] can be very different from reality; as such, engagement and education from trusted sources [19,23] or using realistic AI-based health scenarios [10] are required. This lack of knowledge may also be problematic when considering the process of informed consent for any future AI interventions [24]. Despite this challenge, we identified that patients were generally more trusting of AI than not and a large proportion thought that the benefits outweighed the potential risks.

Patients report that they are more knowledgeable about how the health service in the UK (NHS) uses health data for research than commercial organizations or university researchers. However, most patients would be comfortable sharing anonymized health data with the NHS and university researchers. Both are public institutions, and therefore, this demonstrates the importance of trust when sharing sensitive information. We also identified that patients were less willing to share data with commercial organizations. Privacy fears [7] and anxiety that the transferred data may be used for profit could explain this finding. This was especially the case with news organizations, technology companies, and insurance companies. Our findings add to a downward trend in public trust regarding sharing data with commercial organizations [25], which seems to have changed significantly when compared to historical evidence [26]. This suggests that recent technology scandals such as Cambridge Analytica [27] and media reports of inappropriate sharing of patient data with technology companies [13] have increased public awareness about the potential risks and consequences of data sharing with commercial companies

XSL•FO RenderX

[28]. Governmental guidelines and regulations [29,30] have recently been published to reassure patients that data-driven technology is safe and can maintain privacy, and they provide evidence of what good practice looks like to the industry and commissioners. These findings are similar to a recent systematic review [7], where the conditional nature of support for data sharing was identified. A variety of concerns including data security, privacy, anonymization, and control of data were also raised in this review.

Anonymization of data sets through deidentification is crucial to allow safe storage and sharing of health data while preserving privacy [7]. However, current processes for de identification have proved susceptible to reidentification attacks and the risk of this happening can never be completely eliminated [12]. There is also concerning evidence that even accepted deidentification techniques may not be sufficient to ensure privacy in the face of sophisticated AI algorithms [7]. This is especially concerning as AI research in health care requires large, granular data sets containing sensitive information, which if compromised could cause psychological and reputational harm to patients. Our study demonstrates that patients would be less supportive of data sharing if there was a probability of reidentification. In an attempt to mitigate this concern, the Information Commissioner's Office (ICO), the United Kingdom's independent statutory body for information rights, has issued a code of practice on anonymization [31]. In the United States, the Privacy Rules of the Health Insurance Portability and Accountability Act (HIPAA) provides similar guidance [32]. These guidelines, along with the introduction of the General Data Protection Regulations (GDPR) in Europe and enhanced cybersecurity [33], may allay public fears about reidentification of health data. However, despite these regulations, multiple privacy challenges specific to AI remain and updated ethical and legal frameworks are required to regulate the use of AI in health care [34].

Multivariate analysis revealed some differences in views across participant subgroups. Consistent with previous literature, BAME populations were generally less supportive of data sharing and AI research [10] along with younger age groups and those with lower educational attainment. Training and testing of AI algorithms require diverse data sets that are representative of the local population for which the algorithm will be deployed [35]. The lack of inclusion of minorities in AI data sets has been shown to induce algorithmic bias [36]. Educating BAME communities about the benefits of data sharing is required to help minimize this bias and ensure that AI research is representative of the target population. The differences noted across age groups may be related to the fact that older people may pay more attention to health and medical issues than younger people. There are opportunities to better engage younger people with creative approaches such as through social media, and these should be explored further [37,38].

Notwithstanding the issues outlined above, the majority of respondents in our study who had prior knowledge of AI would support AI research on health care data and imaging in a university setting. However, it is imperative to understand which health data are considered acceptable and unacceptable for AI research by patients. The authors believe that it is important that patients are not simply informed about how health data is used in AI research but are actively involved and consulted with in all aspects of the work. The involvement and guidance of patients and the public will ensure that using AI in health care is transparent, trustworthy, ethical, and socially beneficial.

Limitations

Our results should be interpreted in the context of the limitations related to our study design. This was a cross-sectional questionnaire study that provides a snapshot of patients' views and thoughts, rather than how these may change over time. This is particularly relevant to this study where data was collected 3 years ago because AI research is a rapidly advancing field with an abundance of new research and media articles published regularly. Therefore, it is inevitable that patients' knowledge and viewpoints will change over time. The demographic characteristics of our patients and the fact that patients were recruited from only a UK public hospital may limit the generalizability of the findings. Furthermore, the convenience sampling technique used to approach patients for inclusion in this study signifies that the findings are not likely be generalizable to a wider population that may have no relationship with health services. Cross-sectional studies are also prone to nonresponse bias, which can result in a nonrepresentative sample. Unfortunately, the number of patients who declined to complete the questionnaire was not accurately measured in this study (although approximately 1000 patients were approached); hence, it is difficult to measure the effect of this aspect. There is a risk of selection biases caused by the survey being in English, but we attempted to minimize selection bias by recruiting patients on different days and times and from different areas of the hospitals. Although definitions and clarifications about AI and health data research were provided and we conducted pilot work to simplify the questions, the survey concepts were complex; therefore, some respondents may have not fully understood the information provided.

Conclusions

With increasing research on implementing AI in health care, more attention is given to the public opinion and acceptability of this type of research on health data. This study has demonstrated that there are significant variations in the patients' perception, knowledge and understanding of health data research and AI. There is a need for greater awareness among the public and patients, which can only be achieved by public engagement and debates. This will be instrumental for ensuring the acceptability of AI research and its successful integration into clinical practice in future.

Acknowledgments

Infrastructure support for this research was provided by the NIHR Imperial Biomedical Research Centre. We would like to thank Anna Lawrence-Jones for her help with the recruitment of patients.

Conflicts of Interest

AD is Chair of the Health Security initiative at Flagship Pioneering UK Ltd.

Multimedia Appendix 1
Tables demonstrating the results of the multivariate regression analyses for the survey questions.
[DOCX File , 106 KB - jmir_v23i8e26162_app1.docx]

References

1. Obermeyer Z, Emanuel EJ. Predicting the future-big data, machine learning, and clinical medicine. N Engl J Med 2016 Sep;375(13):1216-1219 [FREE Full text] [doi: 10.1056/NEJMp1606181] [Medline: 27682033]
2. El NI, Murphy M. What is machine learning? In: El NI, Li R, Murphy M, editors. Machine Learning in Radiation Oncology. Switzerland: Springer, Cham; 2015:978-983 URL: https://www.springer.com/gp/book/9783319183046
3. Litjens G, Kooi T, Bejnordi B, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Med Image Anal 2017 Dec;42:60-88 [FREE Full text] [doi: 10.1016/j.media.2017.07.005] [Medline: 28778026]
4. Bluemke D, Moy L, Bredella M, Ertl-Wagner BB, Fowler KJ, Goh VJ, et al. Assessing radiology research on artificial intelligence: a brief guide for authors, reviewers, and readers-from the editorial board. Radiology 2020 Mar;294(3):487-489. [doi: 10.1148/radiol.2019192515] [Medline: 31891322]
5. Beam AL, Kohane IS. Big data and machine learning in health care. JAMA 2018 Apr;319(13):1317-1318. [doi: 10.1001/jama.2017.18391] [Medline: 29532063]
6. Ipsos MORI Social Research Institute. The one-way mirror: public attitudes to commercial access to health data. The one-way mirror: public attitudes to commercial access to health data. 2017. URL: https://wellcome.figshare.com/articles/journal_contribution/The_One-Way_Mirror_Public_attitudes_to_commercial_access_to_health_data/5616448/1 [accessed 2020-11-20]
7. Aitken M, de St Jorre J, Pagliari C, Jepson R, Cunningham-Burley S. Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies. BMC Med Ethics 2016 Nov;17(1):73 [FREE Full text] [doi: 10.1186/s12910-016-0153-x] [Medline: 27832780]
8. Kalkman S, van Delden J, Banerjee A, Tyl B, Mostert M, van Thiel G. Patients' and public views and attitudes towards the sharing of health data for research: a narrative review of the empirical evidence. J Med Ethics 2019 Nov:1-11 [FREE Full text] [doi: 10.1136/medethics-2019-105651] [Medline: 31719155]
9. Ipsos MORI Social Research Institute. Public Views of Machine Learning. 2017. URL: https://royalsociety.org/-/media/policy/projects/machine-learning/publications/public-views-of-machine-learning-ipsos-mori.pdf [accessed 2020-11-20]
10. McCradden MD, Sarker T, Paprica PA. Conditionally positive: a qualitative study of public perceptions about using health data for artificial intelligence research. BMJ Open 2020 Oct;10(10):e039798. [doi: 10.1136/bmjopen-2020-039798]
11. McCradden MD, Baba A, Saha A, Ahmad S, Boparai K, Fadaiefard P, et al. Ethical concerns around use of artificial intelligence in health care research from the perspective of patients with meningioma, caregivers and health care providers: a qualitative study. CMAJ Open 2020 Feb;8(1):E90-E95 [FREE Full text] [doi: 10.9778/cmajo.20190151] [Medline: 32071143]
12. El Emam K, Jonker E, Arbuckle L, Malin B. A systematic review of re-identification attacks on health data. PLoS One 2011;6(12):e28071 [FREE Full text] [doi: 10.1371/journal.pone.0028071] [Medline: 22164229]
13. Hern A. Google DeepMind 1.6m patient record deal 'inappropriate'. The Guardian. 2017. URL: https://www.theguardian.com/technology/2017/may/16/google-deepmind-16m-patient-record-deal-inappropriate-data-guardian-royal-free [accessed 2021-07-29]
14. Martin G, Ghafur S, Kinross J, Hankin C, Darzi A. WannaCry-a year on. BMJ 2018 Jun;361:k2381. [doi: 10.1136/bmj.k2381] [Medline: 29866711]
15. Nelson CA, Pérez-Chada LM, Creadore A, Li SJ, Lo K, Manjaly P, et al. Patient perspectives on the use of artificial intelligence for skin cancer screening: a qualitative study. JAMA Dermatol 2020 May;156(5):501-512. [doi: 10.1001/jamadermatol.2019.5014] [Medline: 32159733]
16. Ongena YP, Haan M, Yakar D, Kwee TC. Patients' views on the implementation of artificial intelligence in radiology: development and validation of a standardized questionnaire. Eur Radiol 2020 Feb;30(2):1033-1040 [FREE Full text] [doi: 10.1007/s00330-019-06486-0] [Medline: 31705254]
17. Stai B, Heller N, McSweeney S, Rickman J, Blake P, Vasdev R, et al. Public perceptions of artificial intelligence and robotics in medicine. J Endourol 2020 Oct;34(10):1041-1048. [doi: 10.1089/end.2020.0137] [Medline: 32611217]
18. Laï MC, Brian M, Mamzer M. Perceptions of artificial intelligence in healthcare: findings from a qualitative survey study among actors in France. J Transl Med 2020 Jan;18(1):14 [FREE Full text] [doi: 10.1186/s12967-019-02204-y] [Medline: 31918710]
19. Balaram B, Greenham T, Leonard J. Artificial intelligence: real public engagement. Royal Society for the encouragement of Arts, Manufactures and Commerce. URL: https://www.thersa.org/globalassets/pdfs/reports/rsa_artificial-intelligence---real-public-engagement.pdf [accessed 2021-07-29]

20. Benjamens S, Dhunnoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. NPJ Digit Med 2020;3:118 [FREE Full text] [doi: 10.1038/s41746-020-00324-0] [Medline: 32984550]
21. Gao S, He L, Chen Y, Li D, Lai K. Public perception of artificial intelligence in medical care: content analysis of social media. J Med Internet Res 2020 Jul;22(7):e16649 [FREE Full text] [doi: 10.2196/16649] [Medline: 32673231]
22. Cave S, Dihal K. Hopes and fears for intelligent machines in fiction and reality. Nat Mach Intell 2019 Feb;1(2):74-78 [FREE Full text] [doi: 10.1038/s42256-019-0020-9]
23. How to stimulate effective public engagement on the ethics of artificial intelligence. Involve. 2019. URL: https://www.involve.org.uk/sites/default/files/field/attachemnt/How%20to%20stimulate%20effective%20public%20debate%20on%20the%20ethics%20of%20artificial%20intelligence%20_0.pdf [accessed 2020-11-21]
24. Schiff D, Borenstein J. How should clinicians communicate with patients about the roles of artificially intelligent team members? AMA J Ethics 2019 Feb;21(2):E138-E145 [FREE Full text] [doi: 10.1001/amajethics.2019.138] [Medline: 30794123]
25. Ghafur S, Van Dael J, Leis M, Darzi A, Sheikh A. Public perceptions on data sharing: key insights from the UK and the USA. The Lancet Digital Health 2020 Sep;2(9):e444-e446. [doi: 10.1016/s2589-7500(20)30161-8]
26. Public attitudes to commercial access to health data. Wellcome Trust. 2016. URL: https://wellcome.ac.uk/sites/default/files/public-attitudes-to-commercial-access-to-health-data-summary-wellcome-mar16.pdf [accessed 2020-11-21]
27. Cadwalladr C, Graham-Harrison E. Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. The Guardian. 2018. URL: https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election [accessed 2021-07-29]
28. Gostin LO, Halabi SF, Wilson K. Health data and privacy in the digital era. JAMA 2018 Jul;320(3):233-234. [doi: 10.1001/jama.2018.8374] [Medline: 29926092]
29. Code of conduct for data-driven health and care technology. Department of Health and Social Care. 2019. 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 2020-11-21]
30. Artificial intelligence: how to get it right. NHSX. 2019. URL: https://www.nhsx.nhs.uk/media/documents/NHSX_AI_report.pdf [accessed 2020-11-21]
31. Rocher L, Hendrickx JM, de Montjoye Y. Estimating the success of re-identifications in incomplete datasets using generative models. Nat Commun 2019 Jul;10(1):3069 [FREE Full text] [doi: 10.1038/s41467-019-10933-3] [Medline: 31337762]
32. Anonymisation: managing data protection risk code of practice. Information Commissioner's Office. 2012. URL: https://ico.org.uk/media/1061/anonymisation-code.pdf [accessed 2020-11-21]
33. Garfinkel SL. De-identification of personal information. National institute of Standards and Technology. 2015. URL: https://www.privacysecurityacademy.com/wp-content/uploads/2021/05/NIST.IR_.8053-De-identification-of-Personal-Information.pdf [accessed 2021-08-03]
34. Ghafur S, Fontana G, Martin G, Grass E, Goodman J, Darzi A. Improving cyber security in the NHS. Imperial College London; Institute of Global Health Innovation. 2019. URL: http://allcatsrgrey.org.uk/wp/download/informatics/Cyber-Security-Ghafur.pdf [accessed 2020-11-20]
35. Price W, Cohen IG. Privacy in the age of medical big data. Nat Med 2019 Jan;25(1):37-43 [FREE Full text] [doi: 10.1038/s41591-018-0272-7] [Medline: 30617331]
36. Barocas S, Selbst AD. Big data's disparate impact. Calif L Rev 2016;104:671.
37. Chouldechova A, Roth A. The frontiers of fairness in machine learning. ArXiv. Preprint posted online on Oct 20, 2018.
38. Wong CA, Merchant RM, Moreno MA. Using social media to engage adolescents and young adults with their health. Healthc (Amst) 2014 Dec;2(4):220-224 [FREE Full text] [doi: 10.1016/j.hjdsi.2014.10.005] [Medline: 25984444]

Abbreviations

AI: artificial intelligence
A-Level: Advanced Level of Secondary Education
BAME: Black, Asian, and minority ethnic
GDPR: General Data Protection Regulations
HIPAA: Privacy Rules of the Health Insurance Portability and Accountability Act
ICO: Information Commissioner’s Office
ML: machine learning
NHS: National Health Service

Edited by R Kukafka; submitted 02.12.20; peer-reviewed by V Patel, B Brumen, S Liu, F Wirth; comments to author 15.03.21; revised version received 04.04.21; accepted 05.07.21; published 26.08.21.

Please cite as:
Aggarwal R, Farag S, Martin G, Ashrafian H, Darzi A
Patient Perceptions on Data Sharing and Applying Artificial Intelligence to Health Care Data: Cross-sectional Survey
J Med Internet Res 2021;23(8):e26162
URL: https://www.jmir.org/2021/8/e26162
doi:10.2196/26162
PMID:34236994

©Ravi Aggarwal, Soma Farag, Guy Martin, Hutan Ashrafian, Ara Darzi. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

A Machine Learning Approach to Passively Informed Prediction of Mental Health Risk in People with Diabetes: Retrospective Case-Control Analysis

Jessica Yu[1], PhD; Carter Chiu[1], PhD; Yajuan Wang[1], PhD; Eldin Dzubur[1], PhD; Wei Lu[1], PhD; Julia Hoffman[1], PsyD

Livongo Health, Inc, Mountain View, CA, United States

Corresponding Author:
Jessica Yu, PhD
Livongo Health, Inc
150 W Evelyn Ave
Ste 150
Mountain View, CA, 94041
United States
Phone: 1 6508048434
Email: jessica.yu@livongo.com

Abstract

Background: Proactive detection of mental health needs among people with diabetes mellitus could facilitate early intervention, improve overall health and quality of life, and reduce individual and societal health and economic burdens. Passive sensing and ecological momentary assessment are relatively newer methods that may be leveraged for such proactive detection.

Objective: The primary aim of this study was to conceptualize, develop, and evaluate a novel machine learning approach for predicting mental health risk in people with diabetes mellitus.

Methods: A retrospective study was designed to develop and evaluate a machine learning model, utilizing data collected from 142,432 individuals with diabetes enrolled in the Livongo for Diabetes program. First, participants' mental health statuses were verified using prescription and medical and pharmacy claims data. Next, four categories of passive sensing signals were extracted from the participants' behavior in the program, including demographics and glucometer, coaching, and event data. Data sets were then assembled to create participant-period instances, and descriptive analyses were conducted to understand the correlation between mental health status and passive sensing signals. Passive sensing signals were then entered into the model to train and test its performance. The model was evaluated based on seven measures: sensitivity, specificity, precision, area under the curve, F_1 score, accuracy, and confusion matrix. SHapley Additive exPlanations (SHAP) values were computed to determine the importance of individual signals.

Results: In the training (and validation) and three subsequent test sets, the model achieved a confidence score greater than 0.5 for sensitivity, specificity, area under the curve, and accuracy. Signals identified as important by SHAP values included demographics such as race and gender, participant's emotional state during blood glucose checks, time of day of blood glucose checks, blood glucose values, and interaction with the Livongo mobile app and web platform.

Conclusions: Results of this study demonstrate the utility of a passively informed mental health risk algorithm and invite further exploration to identify additional signals and determine when and where such algorithms should be deployed.

(J Med Internet Res 2021;23(8):e27709) doi:10.2196/27709

KEYWORDS

diabetes mellitus; mental health; risk detection; passive sensing; ecological momentary assessment; machine learning

Introduction

In the United States, 34.2 million people are affected by diabetes mellitus [1]. Approximately 25% of those living with diabetes experience significant depressive symptoms, and up to 40% experience generalized anxiety disorder (GAD) [1-3]. Individuals with diabetes and mental health challenges have been found to be less adherent to diabetes treatment recommendations, including diet, exercise, medication use, glucose monitoring, and medical appointments, and they are at a greater risk for adverse medical outcomes [1]. Health care costs for those with comorbid diabetes and mental health have been estimated to be US $4 billion to $9 billion greater than for those without these conditions [4]. However, proactive detection of mental health needs of people with diabetes could facilitate early intervention, thereby improving their overall health and quality of life and reducing the health and economic burdens placed on this population and the health care system as a whole.

Despite recommendations by the American Diabetes Association and the United States Preventive Services Task Force to routinely evaluate people with diabetes for their mental health needs, only 25% to 50% of people with diabetes who have depression receive a mental health diagnosis and intervention [5,6]. This gap in receiving care is a result of a shortage of mental health professionals available to offer assessment and intervention, a lack of mental health knowledge among primary care providers who most often care for patients with diabetes, and limited access to mental health screening tools in health care practices offering services to these patients [6]. Newer methods such as passive sensing and ecological momentary assessments (EMAs) provide a more scalable, less effort- and time-intensive approach to information gathering and assessment. Passive sensing refers to the capture of data about a person without any extra effort on their part [7]. EMA refers to the repeated sampling of an individual's behavior in real time within their natural environment [8]. Both methods can be integrated into or with devices and services that people with diabetes already utilize in their daily lives, such as blood glucose meters, smartphones, and health coaching platforms to enable the collection and processing of data in real time and to provide context for real-time interventions [7].

Although passive sensing and EMA have previously been examined in the general population, limited studies have focused on the detection of mental health needs among the diabetes population outside of using smartphones as data warehouses, relying on accelerometer, GPS, ambient light sensors, and call log data [9,10]. Moreover, no known study to date has attempted to detect mental health concerns in people with diabetes by using blood glucose meters despite the fact that individuals living with diabetes are encouraged to engage with these devices at regular intervals, blood glucose monitoring has been found to be correlated with psychological effects, and engagement with these devices and testing blood glucose levels are known to be associated with mood or stress [11-13]. Further, blood glucose meter data can be paired with data from other sources for a robust view of a person's behavioral and emotional profile. The primary aim of this study was to conceptualize, develop, and evaluate a novel approach using passive sensing for predicting mental health risk in people with diabetes.

Methods

Study Design

A multidisciplinary team of experts in data science, machine learning, and clinical and experimental psychology collaborated in the development of a machine learning model for detecting potential mental health risk from passive sensing signals that was both clinically relevant and statistically rigorous. A retrospective analysis was performed to evaluate the machine learning model for detecting potential mental health risk from passive sensing signals leveraging data collected during participants' engagement in the Livongo for Diabetes program [14].

Livongo for Diabetes

The Livongo for Diabetes program is a digital remote program for the management of chronic condition focused on empowering members by providing education and tools to self-manage their diabetes through mobile technology. The program offers members (1) a cellular-enabled, two-way messaging device that measures blood glucose and delivers personalized insights; (2) free, unlimited blood glucose test strips; (3) real-time support from diabetes response specialists available 24 hours a day, 7 days a week, 365 days a year; and (4) access to certified diabetes care and education specialists for support and goal setting. Further details on the Livongo for Diabetes program and its efficacy in improving diabetes-related outcomes are available in the literature [15-17].

Study Participants

Study participants were defined as those individuals enrolled in the Livongo for Diabetes Program between January 1, 2018, and February 28, 2020, who used their blood glucose meter at least once (N=142,432). Approval was granted by the Aspire Independent Review Board (#520160099). All participants provided consent to participate, and guidelines outlined in the Declaration of Helsinki were followed.

Study Procedure

The mental health status of each participant was verified through available data from two sources that included data on medications prescribed to and filled by participants and mental health–related interventions. Next, passive sensing signals were extracted from participants' behaviors as they interacted with Livongo's blood glucose meter, mobile app, web portal, and coaching feature. Then, data sets were assembled by aggregating these signals per participant over various periods, creating *participant-period* instances. Descriptive analyses were conducted to understand the correlation between the signals and mental health status. Finally, demographic information and passive sensing signals were entered into the model, training it to understand the relationships between these signals and the participants' mental health status.

Study Population Identification

Identification of population *cases* and *controls* with respect to mental health conditions—which, in the context of labeling data for model training, we refer to as *ground truth*—was performed utilizing two data sources: (1) claims data and (2) medication prescription data. Claims data contained information on medications indicated for mental health conditions that were prescribed to and filled by participants, as well as mental health–related assessments and interventions. Based on data availability and right to use, 6.1% (8633/142,432) of the study participants had claims data, which provided a diverse way to identify their mental health needs through diagnoses, procedures, and prescriptions. Medication prescription data contained only information on mental health–related medications that were prescribed to and filled by participants. Medication prescription data were available for the entire study population and used to identify cases for the remaining participants. Participant demographic characteristics, which were used to evaluate signals correlated with mental health in the study population, are summarized in Table 1.

Table 1. Participant demographics and characteristics at the time of enrollment (N=142,432).

Characteristic	Value
Age in years, mean (SD)	54.8 (12.4)
Gender, n (%)	
Female	68,968 (48.4)
Male	73,147 (51.6)
Other	317 (0.22)
Ethnicity, n (%)	
Hispanic	12,809 (9.0)
Non-Hispanic	86,116 (60.5)
Unknown	43,507 (30.6)
Race, n (%)	
Caucasian	66,551 (46.7)
Black or African American	14,702 (10.3)
Asian	8199 (5.76)
Pacific Islander	468 (0.33)
American Indian	725 (0.51)
Other	6,588 (4.63)
Unknown	45,199 (31.7)
Diabetes type, n (%)	
Type 1	14,360 (10.1)
Type 2	126,369 (88.7)
Unknown	1603 (1.2)
Years since diagnosis, mean (SD)	8.27 (8.1)
First reported A_{1c}, mean (SD)	7.51 (1.7)
Insulin use, n (%)	
Yes	39,153 (27.5)
No	102,622 (72.1)
Unknown	657 (0.5)

Passive Sensing Signals

All data utilized in the study were collected in the course of how participants naturally engaged with the Livongo for Diabetes Program. That is, no data were collected solely for study purposes. We identified various data sources potentially useful to detect mental health risk behaviors. From these data sources, we extracted 83 individual signals that can be broadly classified into the following four categories. Note that individual signal names are withheld to protect proprietary information.

Demographics

Demographic factors such as age, gender, ethnicity, and race have been shown to be related to mental health [18]. Therefore, we included participants' demographic data into the model.

Glucometer Data

The Livongo blood glucose meter is the most frequent interaction point for participants of the Livongo for Diabetes Program. Low rates of blood glucose monitoring [2] and poorer blood glucose control [19] have been linked to depression among those with diabetes; and depression, anxiety, and stress symptoms are greater among people with diabetes than those without [20]. Differences in device usage is particularly informative of conditions such as depression when examining usage time of day [10] and weekday [21]. Therefore, the key metrics derived from glucometer usage included the number of times blood glucose was checked; blood glucose levels; and variations, responses to questions to assess context such as current emotional state, and time of the day and day of week when the reading was taken.

Coaching Data

In the Livongo for Diabetes Program, Livongo coaches contact individuals under certain conditions. Numerous studies have affirmed relationships between sociability and mental health. Fewer calls and fewer incoming texts have been linked to depression [22], whereas frequency and duration of conversations have been shown to be useful in evaluation of bipolar disorder [23]. Coaching data can serve as a proxy for sociability, for which successful or failed contacts and time spent interacting can be used to glean valuable insights.

Event Data

In addition to the blood glucose meter, individuals enrolled in the Livongo for Diabetes Program interact with multiple platforms during program participation, including the Livongo mobile app and web portal. In following the motivations behind utilizing glucometer and coaching data, we collected frequency, duration, interactivity, and consistency of interaction sessions, as well as the time of day and day of week information associated with the use of the mobile app and web portal. In addition, we tracked voluntary report sharing with friends and family as well as interactions with pop-up reminders.

Statistical Analyses

We conducted correlation analyses using Pearson *r* to preliminarily gauge the strength of the relationship between each extracted signal and the presence of mental health conditions. Among the most highly correlated signals, demographics were well represented, including gender (male: r=–0.156, P<.001; female: r=0.155, P<.001), race (Asian: r=–0.104, P<.001; White: r=0.101, P<.001; Black: r=–0.041, P<001), BMI (r=0.086, P<.001), and smoking status (active smoker: r=0.047, P<.001). With regard to glucometer, coaching, and event data, responses on current emotional state from participants during blood glucose checks indicating wellness (r=–0.108, P<.001) or unwellness (r=0.121, P<.001) were most strongly correlated. Greater frequency and consistency of interactions with services were negatively associated with mental health conditions. In addition, we found that greater variation in blood glucose values, as measured by SD, were positively correlated with mental health conditions (r=0.041, P<.001).

Outcome Data

When quantifying the performance of a prospective model for identifying mental health risk, it is important to consider a variety of perspectives. Metrics such as total accuracy are inadequate if used alone because they can hide model deficiencies on imbalanced data. Consider these three questions, which cannot be answered with total accuracy alone but are of particular importance in a diagnostic setting:

1. How often are those with mental health needs (cases) correctly identified?
2. How often are those without mental health needs (controls) correctly identified?
3. How often are those with predicted mental health needs truly cases?

The following seven measures commonly used in machine learning model evaluation [24] were selected to address the above questions and beyond, enabling a holistic view of model performance:

1. Sensitivity or recall, which addresses question 1
2. Specificity, which addresses question 2
3. Precision, which addresses question 3
4. Area under the curve (AUC), defined as the area under the receiver operating characteristic curve—an important measure quantifying the model’s capacity to differentiate *cases* and *controls* (1=ideal performance, 0.5=random prediction)
5. F_1 score, defined as the harmonic mean of the precision and recall
6. Accuracy, defined as the proportion of instances correctly classified
7. The confusion matrix, which depicts the number of correctly and incorrectly identified cases and controls

Model Development

To develop our machine learning model, we then had to divide the study population into two segments. The first segment, termed the training and validation set, enabled the model to learn. The second segment, termed the test sets, was held separate from model training and used to evaluate the model’s ability to generalize to unseen data. For training, we used a time-interval slice of the population consisting of 124,322 participants (ie, 87% of the study population) who had activated their blood glucose meters in 2018 or 2019, with passive sensing signal data collected in that timeframe. For testing, we defined three distinct test sets designed to comprehensively evaluate model performance. The first two test sets used medication prescription or refill data as their sources of ground truth, whereas the third test set used medical and pharmacy claims data as its source of ground truth.

1. The first test set (test set 1) consisted of the same participant subset as the training data, but with signal data collected in the first two months of 2020. This data subset evaluated model prediction capability on previously seen participants (93,155/142,432, 65.4%).
2. The second test set (test set 2) consisted of participants activated in the first two months of 2020 and the associated signal data. This test set evaluated model prediction capability on new, unseen participants (9477/142,432, 6.7%).
3. The third test set (test set 3) utilized the claims data by identifying participants activated in 2018, 2019, or the first two months of 2020. This final data subset evaluated prediction capability with regard to unseen participants, with mental health needs identified through more diverse sources beyond prescriptions only (8633/142,432, 6.1%).

A visual summary of the data subsets is presented in Figure 1, and the specific numeric breakdowns are described in Table 2. Furthermore, the demographic information for each subset is detailed in Multimedia Appendix 1.

To increase the model utility, passive signals were aggregated during a certain period (participant-period) and presented to the model for prediction of mental health risk. We defined the participant-period as an instance. In this study, an aggregation window of 4 weeks was selected to optimize data availability. Furthermore, two additional conditions were applied to filter out ineligible participant-period instances: (1) instances before a participant had participated in the Livongo Program and (2) instances of extended inactivity, defined as 30 or more days without any interaction with the Livongo for Diabetes Program.

The rationale for the first criterion is trivial. The second eligibility condition reflects the reasoning that a model for identifying mental health needs from passive signals should only be employed when a signal is present, specifically signals where missingness cannot be assumed to be zero (eg, blood glucose values). Thus, our model should only be trained and evaluated on complete instances. Table 3 demonstrates the distribution of eligible instances among the data subsets. Note the class imbalance, with control instances represented at roughly a 2:1 ratio over case instances in each data subset.

Figure 1. Source and date ranges of data subsets defined in this study.

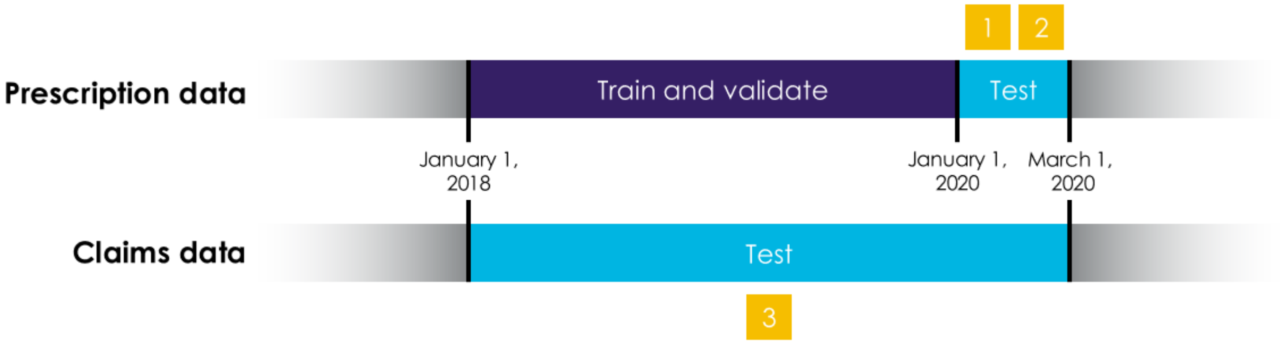


Table 2. Number of participant cases and controls for each data subset.

Data subset	Cases, n (%)	Controls, n (%)	Subset total, n
Training or validation set	42,481 (34.2)	81,841 (65.8)	124,322
Test set 1	31,251 (33.5)	61,904 (66.5)	93,155
Test set 2	2919 (30.8)	6558 (69.2)	9477
Test set 3	3358 (38.9)	5275 (61.1)	8633
Total unique	48,758 (34.2)	93,674 (65.8)	142,432

Table 3. Number of participant-period cases and controls for each data subset.

Data subset	Cases, n (%)	Controls, n (%)	Subset total, n
Training and validation set	287,311 (34.3)	549,183 (65.7)	836,494
Test set 1	54,709 (33.3)	109,550 (66.7)	164,259
Test set 2	2953 (30.8)	6640 (69.2)	9593
Test set 3	34,190 (40.1)	51,150 (59.9)	85,340
Total	379,163 (34.6)	716,523 (65.4)	1,095,686

A machine learning model was enlisted to capture the relationship between input activity features and exhibited mental health needs. The core component of our approach was the training of LightGBM [25] gradient tree boosting models on random subsets of the training data. This approach addressed the class imbalance; we undersampled the training control instances by random undersampling, thus reducing the number of control instances to equal the number of case instances. Because this technique reduced the number of control instances by roughly half, we saw the opportunity to train multiple models on multiple random subsets. This strategy enabled us to fully utilize the entire training data set, with each model training on a differing perspective of the data. We utilized soft voting to obtain an output prediction for a given instance, meaning the outputs of each constituent model—a value from 0 to 1 interpretable as the confidence that an instance is a case—were averaged to obtain a single aggregate confidence score.

Our final devised model consists of an ensemble of 10 LightGBM models. During our model selection process, we evaluated multiple other classes of machine learning models, including logistic regression, random forests, and neural networks. We also experimented with other flavors of gradient boosting, including XGBoost and CatBoost, and overall, LightGBM yielded the highest performance for our training task. To tune each constituent model, we used automated hyperparameter tuning enabled by the hyperopt [26] Python library with 5-fold cross-validation on the training set. Following this training procedure, we evaluated the model on the three held-out test sets to assess model performance.

Results

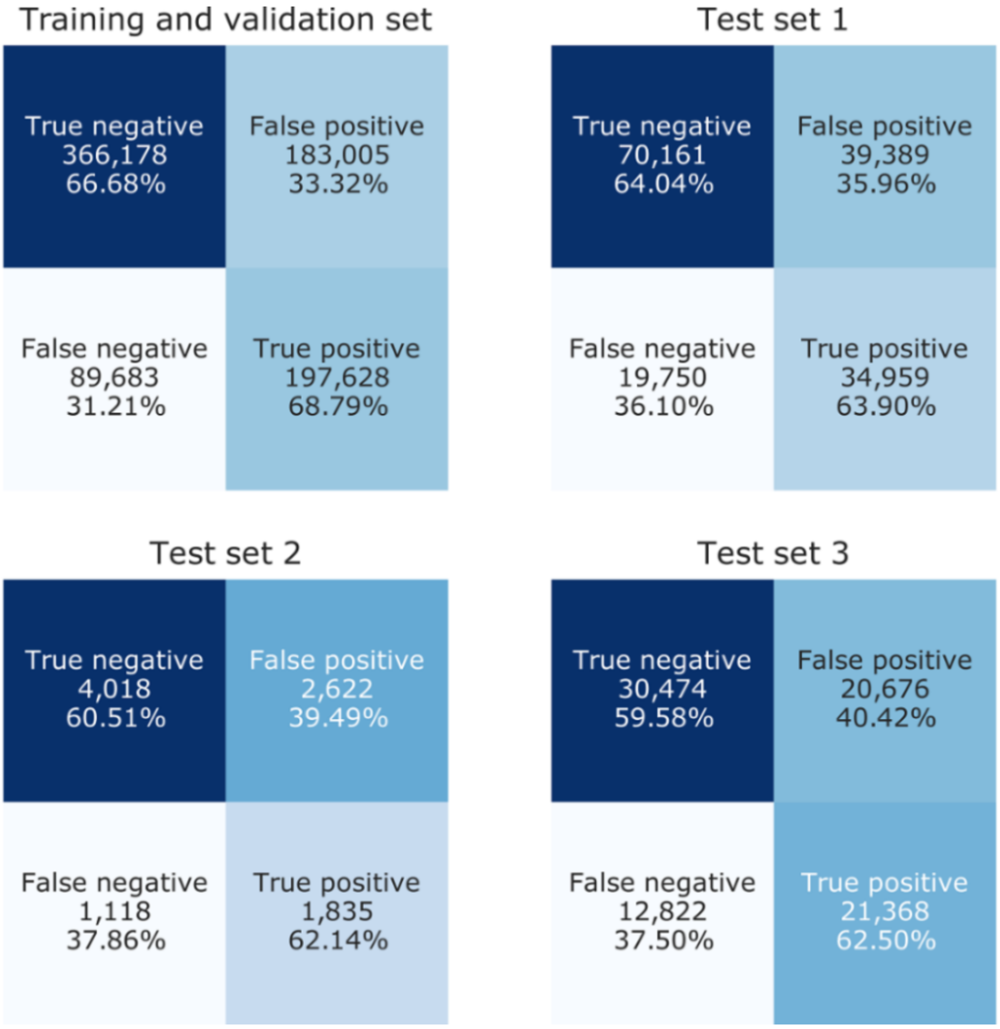
The results of our model's performance on each of the previously described data subsets are presented in Table 4. In addition, the associated confusion matrices are shown in Figure 2, with both counts and percentages (normalized by class support size) depicted. Across all three test sets, the vast majority of metrics exceeded 0.5. Notably, we achieved an AUC of nearly 0.7 on the first test set and exceeded 0.65 across all three sets. The metrics for which the model fell short of the 0.5 mark were precision in the first and second test sets and the F_1 score for the second test set. However, it is important to note that owing to class imbalance, 0.5 would not be the theoretical precision or the F_1 score yielded by random prediction. Rather, the precision obtained by random prediction would be the proportion of cases, with the F_1 score affected commensurately. In our case, baseline precision would be approximately 0.3331 for the first test set and 0.3078 for the second, both of which were well outperformed by the reported precisions of 0.4702 and 0.4164. Likewise, the baseline theoretical F_1 score of 0.3810 was greatly outperformed by the reported 0.4953 on the second test set. Concerning improvement over random prediction overall, our model produced an approximately 14-point gain in precision for the first test set and an approximately 10-point gain for the second and third sets, whereas recall improved by 14 and 12 points and AUC improved by 20 and 16 points, respectively. These results demonstrate a respectable, generalizable performance and are an encouraging advancement towards practical passive mental health risk assessment at scale.

Table 4. Performance metrics for each data subset.

Data subset	Sensitivity	Specificity	Precision	AUC[a]	F_1 Score	Accuracy
Training and validation set	0.688	0.667	0.519	0.745	0.592	0.674
Test set 1	0.639	0.640	0.470	0.696	0.542	0.640
Test set 2	0.621	0.605	0.412	0.658	0.495	0.610
Test set 3	0.625	0.596	0.508	0.656	0.561	0.608

[a] AUC: area under the curve.

Figure 2. Confusion matrices for each data subset.



Our model results also provided the opportunity to gain further insight into the utility of the passive sensing signals. For this purpose, we used the SHapley Additive exPlanations (SHAP) method [27] to compute feature importance, allowing us to quantify the average contribution of each signal to the model. These values are presented in Table 5. Our findings closely mirrored the insights from our correlation analysis. Demographics ranked highly among signals, with gender and race identified as the two most relevant factors. Among passive sensing signals, responses indicating an emotional state of wellness or unwellness during blood glucose checks were deemed the most important. Interaction frequency and consistency were also considered valuable according to SHAP. The model took into particular consideration blood glucose checks based on the times of day and the proportion of days they were performed. Mean and SD values of blood glucose levels also appeared in the top quartile of signals.

Table 5. Feature importance as measured by mean absolute SHapley Additive exPlanations (SHAP) values, interpreted as the average impact on model output (log-odds) magnitude. Note that aliases based on the signal category are depicted in lieu of the signals to protect proprietary information.

Passive signal	Mean absolute SHAP value
Demographics 1	0.152
Demographics 2	0.089
Glucometer data 1	0.078
Demographics 3	0.040
Glucometer data 2	0.039
Demographics 4	0.037
Demographics 5	0.037
Event data 1	0.022
Demographics 6	0.018
Event data 2	0.017
Glucometer data 3	0.014
Glucometer data 4	0.013
Glucometer data 5	0.012
Glucometer data 6	0.012
Glucometer data 7	0.011
Demographics 7	0.011
Glucometer data 8	0.011
Glucometer data 9	0.010
Glucometer data 10	0.008
Glucometer data 11	0.008
Glucometer data 12	0.008
Glucometer data 13	0.007
Event data 3	0.006
Event data 4	0.006
Glucometer data 14	0.006
Glucometer data 15	0.006
Event data 5	0.006
Event data 6	0.005
Glucometer data 16	0.005
Event data 7	0.005
Event data 8	0.004
Event data 9	0.004
Event data 10	0.004
Glucometer data 17	0.004
Event data 11	0.003
Glucometer data 18	0.003
Coaching data 1	0.003
Glucometer data 19	0.002
Event data 12	0.002
Event data 13	0.002
Coaching data 2	0.002

Passive signal	Mean absolute SHAP value
Event data 14	0.002
Event data 15	0.002
Glucometer data 20	0.002
Event data 16	0.002
Coaching data 3	0.001
Event data 17	0.001
Event data 18	0.001
Coaching data 4	0.001
Event data 19	0.001
Event data 20	<0.001
Event data 21	<0.001
Event data 22	<0.001
Event data 23	<0.001
Coaching data 5	<0.001
Event data 24	<0.001
Event data 25	<0.001
Event data 26	<0.001
Event data 27	<0.001
Event data 28	<0.001
Coaching data 6	<0.001
Coaching data 7	<0.001
Event data 29	<0.001
Event data 30	<0.001
Event data 31	<0.001
Event data 32	<0.001
Event data 33	<0.001
Event data 34	<0.001
Event data 35	<0.0001
Event data 36	<0.0001
Event data 37	<0.0001
Event data 38	<0.0001
Event data 39	<0.0001
Event data 40	<0.0001
Event data 41	<0.0001
Event data 42	<0.0001
Event data 43	<0.0001
Event data 44	<0.0001
Coaching data 8	<0.0001
Coaching data 9	0
Coaching data 10	0
Coaching data 11	0
Event data 45	0

Discussion

In this study, we found that a machine learning approach using passive sensing signals that included data on participant demographics, blood glucose meter use, interaction with diabetes coaches as a proxy for sociability, and engagement with the Livongo for Diabetes Program demonstrated utility in predicting mental health risk among people with diabetes.

The results of our approach invite further exploration and expansion. It is well understood that smartphones can be viewed as vehicles for passive data collection and help identify digital phenotypes of mental health disorders, as shown previously [28]. However, it is time to move beyond focusing on smartphones as the only devices that enable passive sensing and EMA and view other devices and services that people with diabetes must use for their self-management as robust data warehouses. In this particular study, participants who enrolled in the Livongo for Diabetes Program had access to a Bluetooth-enabled blood glucose meter for measuring their blood glucose levels, the Livongo mobile app and web platform for tracking food intake and physical activity as well as receiving health reminders, and Livongo coaches for coaching for diabetes self-management. Each device and service offered valuable data to input in our model. The blood glucose meters provided access to the participants' behavioral, emotional, and physiological data, such as how they were feeling at the time of measuring their blood glucose level and the reading itself. The Livongo mobile app and web platform enabled us to understand when participants were awake, using their smartphones, and engaged in a health-related activity. Coaching allowed us to understand whether participants were actively communicating with others. Together, these different data sources enabled us to create a data set that combined behavioral, emotional, and physiological factors into a holistic predictive algorithm. Although these particular devices and services are unique to Livongo members, there are ways to obtain similar data in the real world. For example, several commercially available wireless and Bluetooth-enabled blood glucose meters connect to mobile apps that enable people with diabetes mellitus to track and receive feedback on their blood glucose levels and share data with others. Such meters and associated apps host behavioral, communication, and physiological data similar to what we used in our model. There is also a plethora of health-related apps that enable individuals to track their food intake, physical activity, mood, sleep, and other health signals. These apps host additional behavioral and emotional data similar to what we used in our model.

Identifying potential mental health risk from passively collected signals is undoubtedly not a simple task, and our study has some limitations. First, because we limited our extracted signals to interactions with Livongo devices and applications, we did not have access to certain passive signals shown to be predictive in previous studies, such as mobile device accelerometer, ambient light sensor, or GPS data. As a result, our model was given a somewhat restricted view of a member's activity and sociability patterns. Second, we had access to a limited volume of medical and pharmacy claims data, which made it difficult to utilize the data on their own. However, both limitations could also be seen as strengths of our study. Our inability to access previously studied passive signals afforded us an opportunity to examine new signals from devices and services that are unique to people with diabetes. It may also be more acceptable from a privacy perspective. Furthermore, we had limited claims data; nevertheless, the data enabled us to confidently label participants as *cases* versus *controls*. Finally, a major strength of our study was the fact that, by design, no data were collected with active participant input for the express purpose of detecting mental health risk. In that regard, our proposed approach makes real-world deployment more readily feasible, in contrast to other studies of passive sensing and EMA for mental health, which required active participant participation and extensive sensor infrastructure.

In sum, our model is a bold step toward detecting potential mental health risk passively and autonomously. In its nascent stage, we recommend integrating such a model with existing systems and services, while continuing to improve the quality and completeness of care that can be offered to those dealing with mental health needs.

Conflicts of Interest

The authors own stock in Teladoc Health and are employees of Livongo Health, Inc.

Multimedia Appendix 1

Demographics distribution across various data sets.

[DOCX File , 20 KB - jmir_v23i8e27709_app1.docx]

References

1. National Diabetes Statistics Report, 2020. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/diabetes/data/statistics-report/index.html [accessed 2020-12-01]
2. Holt RIG, de Groot M, Golden SH. Diabetes and Depression. Curr Diab Rep 2014 Apr 18;14(6):1-17. [doi: 10.1007/s11892-014-0491-3]
3. Grigsby AB, Anderson RJ, Freedland KE, Clouse RE, Lustman PJ. Prevalence of anxiety in adults with diabetes: a systematic review. J Psychosom Res 2002 Dec;53(6):1053-1060. [doi: 10.1016/s0022-3999(02)00417-8] [Medline: 12479986]

4. Egede LE, Walker RJ, Bishu K, Dismuke CE. Trends in costs of depression in adults with diabetes in the united states: Medical Expenditure Panel Survey, 2004-2011. J Gen Intern Med 2016 Jun;31(6):615-622. [doi: 10.1007/s11606-016-3650-1] [Medline: 26969312]
5. Diabetes and mental health. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/diabetes/managing/mental-health.html [accessed 2020-12-01]
6. Owens-Gary MD, Zhang X, Jawanda S, Bullard KM, Allweiss P, Smith BD. The importance of addressing depression and diabetes distress in adults with type 2 diabetes. J Gen Intern Med 2019 Feb;34(2):320-324 [FREE Full text] [doi: 10.1007/s11606-018-4705-2] [Medline: 30350030]
7. Cornet VP, Holden RJ. Systematic review of smartphone-based passive sensing for health and wellbeing. J Biomed Inform 2018 Jan;77:120-132 [FREE Full text] [doi: 10.1016/j.jbi.2017.12.008] [Medline: 29248628]
8. 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]
9. Marsch LA. Digital health data-driven approaches to understand human behavior. Neuropsychopharmacology 2021 Jan;46(1):191-196 [FREE Full text] [doi: 10.1038/s41386-020-0761-5] [Medline: 32653896]
10. Sarda A, Munuswamy S, Sarda S, Subramanian V. Using passive smartphone sensing for improved risk stratification of patients with depression and diabetes: cross-sectional observational study. JMIR Mhealth Uhealth 2019 Jan 29;7(1):e11041 [FREE Full text] [doi: 10.2196/11041] [Medline: 30694197]
11. AADE7 Self-Care Behaviors. Association of Diabetes Care & Education Specialists. URL: https://www.diabeteseducator.org/living-with-diabetes/aade7-self-care-behaviors [accessed 2020-12-01]
12. Patton SR, Clements MA. Psychological reactions associated with continuous glucose monitoring in youth. J Diabetes Sci Technol 2016 May;10(3):656-661 [FREE Full text] [doi: 10.1177/1932296816638109] [Medline: 26969141]
13. Fisher L, Polonsky W, Parkin CG, Jelsovsky Z, Amstutz L, Wagner RS. The impact of blood glucose monitoring on depression and distress in insulin-naïve patients with type 2 diabetes. Curr Med Res Opin 2011 Nov;27 Suppl 3:39-46. [doi: 10.1185/03007995.2011.619176] [Medline: 21916532]
14. Livongo.com. URL: https://www.livongo.com/ [accessed 2021-06-24]
15. Downing J, Bollyky J, Schneider J. Use of a connected glucose meter and certified diabetes educator coaching to decrease the likelihood of abnormal blood glucose excursions: the Livongo for Diabetes Program. J Med Internet Res 2017 Jul 11;19(7):e234 [FREE Full text] [doi: 10.2196/jmir.6659] [Medline: 28698167]
16. Bollyky JB, Bravata D, Yang J, Williamson M, Schneider J. Remote lifestyle coaching plus a connected glucose meter with certified diabetes educator support improves glucose and weight loss for people with type 2 diabetes. J Diabetes Res 2018;2018:3961730 [FREE Full text] [doi: 10.1155/2018/3961730] [Medline: 29888288]
17. 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 Mar 08:1-9. [doi: 10.1080/13696998.2019.1609483] [Medline: 31012392]
18. Fleury M, Grenier G, Bamvita J, Perreault M, Jean-Caron. Typology of adults diagnosed with mental disorders based on socio-demographics and clinical and service use characteristics. BMC Psychiatry 2011 Apr 20;11(67):1-11. [doi: 10.1186/1471-244x-11-67]
19. Lustman PJ, Anderson RJ, Freedland KE, de Groot M, Carney RM, Clouse RE. Depression and poor glycemic control: a meta-analytic review of the literature. Diabetes Care 2000 Jul;23(7):934-942 [FREE Full text] [doi: 10.2337/diacare.23.7.934] [Medline: 10895843]
20. Krishna P. Depression, anxiety, and stress levels in patients with type 2 diabetes mellitus. Natl J Physiol Pharm Pharmacol 2018;8(9):1570-1572. [doi: 10.5455/njppp.2018.8.0929117092018]
21. Saeb S, Lattie EG, Schueller SM, Kording KP, Mohr DC. The relationship between mobile phone location sensor data and depressive symptom severity. PeerJ 2016;4:e2537 [FREE Full text] [doi: 10.7717/peerj.2537] [Medline: 28344895]
22. Ghandeharioun A, Fedor S, Sangermano L. Objective assessment of depressive symptoms with machine learning and wearable sensors data. 2017 Presented at: Seventh International Conference on Affective Computing and Intelligent Interaction; 2017; San Antonio, TX p. 325-332. [doi: 10.1109/acii.2017.8273620]
23. Abdullah S, Matthews M, Frank E, Doherty G, Gay G, Choudhury T. Automatic detection of social rhythms in bipolar disorder. J Am Med Inform Assoc 2016 May;23(3):538-543. [doi: 10.1093/jamia/ocv200] [Medline: 26977102]
24. Japkowicz N, Shah M. Evaluating learning algorithms: a classification perspective. New York, NY: Cambridge University Press; 2011.
25. Ke G, Meng Q, Finley T. LightGBM: a highly efficient gradient boosting decision tree. 2017 Presented at: 31st Conference on Neural Information Processing Systems; 2017; Long Beach, CA.
26. Bergstra J, Yamins D, Cox D. Making a science of model search: hyperparameter optimization in hundreds of dimensions for vision architectures. 2013 Presented at: Proceedings of the 30th International Conference on Machine Learning; 2013; Atlanta, GA.
27. Lundberg S, Lee S. A unified approach to interpreting model predictions. 2017 Presented at: 31st Conference on Neural Information Processing Systems; 2017; Long Beach, CA.

28. Zulueta J, Piscitello A, Rasic M, Easter R, Babu P, Langenecker SA, et al. Predicting mood disturbance severity with mobile phone keystroke metadata: a BiAffect digital phenotyping study. J Med Internet Res 2018 Jul 20;20(7):e241 [FREE Full text] [doi: 10.2196/jmir.9775] [Medline: 30030209]

Abbreviations

AUC: area under the curve
EMA: ecological momentary assessment
GAD: generalized anxiety disorder
SHAP: SHapley Additive exPlanations

Edited by R Kukafka; submitted 04.02.21; peer-reviewed by L Shen, P Nelson, E Sükei; comments to author 18.03.21; revised version received 22.04.21; accepted 24.05.21; published 27.08.21.

Please cite as:
Yu J, Chiu C, Wang Y, Dzubur E, Lu W, Hoffman J
A Machine Learning Approach to Passively Informed Prediction of Mental Health Risk in People with Diabetes: Retrospective Case-Control Analysis
J Med Internet Res 2021;23(8):e27709
URL: https://www.jmir.org/2021/8/e27709
doi:10.2196/27709
PMID:34448707

©Jessica Yu, Carter Chiu, Yajuan Wang, Eldin Dzubur, Wei Lu, Julia Hoffman. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Image Processing for Public Health Surveillance of Tobacco Point-of-Sale Advertising: Machine Learning–Based Methodology

Ned English[1], MSc; Andrew Anesetti-Rothermel[2], MPH, PhD; Chang Zhao[1], PhD; Andrew Latterner[1], BA; Adam F Benson[3], MSPH; Peter Herman[1], MA; Sherry Emery[1], PhD; Jordan Schneider[3], MPH; Shyanika W Rose[4], MA, MPH, PhD; Minal Patel[3], MPH, PhD; Barbara A Schillo[3], PhD

[1]NORC at the University of Chicago, Chicago, IL, United States

[2]Center for Tobacco Products, US Food and Drug Administration, Silver Spring, MD, United States

[3]Truth Initiative Schroeder Institute, Washington, DC, United States

[4]Department of Behavioral Science, College of Medicine and the Center for Health Equity Transformation, University of Kentucky, Lexington, KY, United States

Corresponding Author:
Ned English, MSc
NORC at the University of Chicago
55 E Monroe St
Ste 3100
Chicago, IL, 60603
United States
Phone: 1 3127594010
Fax: 1 3127594000
Email: english-ned@norc.org

Abstract

Background: With a rapidly evolving tobacco retail environment, it is increasingly necessary to understand the point-of-sale (POS) advertising environment as part of tobacco surveillance and control. Advances in machine learning and image processing suggest the ability for more efficient and nuanced data capture than previously available.

Objective: The study aims to use machine learning algorithms to discover the presence of tobacco advertising in photographs of tobacco POS advertising and their location in the photograph.

Methods: We first collected images of the interiors of tobacco retailers in West Virginia and the District of Columbia during 2016 and 2018. The clearest photographs were selected and used to create a training and test data set. We then used a pretrained image classification network model, Inception V3, to discover the presence of tobacco logos and a unified object detection system, You Only Look Once V3, to identify logo locations.

Results: Our model was successful in identifying the presence of advertising within images, with a classification accuracy of over 75% for 8 of the 42 brands. Discovering the location of logos within a given photograph was more challenging because of the relatively small training data set, resulting in a mean average precision score of 0.72 and an intersection over union score of 0.62.

Conclusions: Our research provides preliminary evidence for a novel methodological approach that tobacco researchers and other public health practitioners can apply in the collection and processing of data for tobacco or other POS surveillance efforts. The resulting surveillance information can inform policy adoption, implementation, and enforcement. Limitations notwithstanding, our analysis shows the promise of using machine learning as part of a suite of tools to understand the tobacco retail environment, make policy recommendations, and design public health interventions at the municipal or other jurisdictional scale.

(J Med Internet Res 2021;23(8):e24408) doi:10.2196/24408

KEYWORDS

machine learning; image classification; convolutional neural network; object detection; crowdsourcing; tobacco point of sale; public health surveillance

Introduction

Background

Tobacco point-of-sale (POS) advertising, consisting of signs, displays, and other promotional materials, is considered a very deliberate and effective marketing strategy [1]. The 1998 Master Settlement Agreement restricted tobacco advertising in general, raising the importance of POS advertising as one of the only remaining channels tobacco companies could use to directly reach consumers. In 2018, POS advertising represented the largest category of advertising expenditure for cigarette manufacturers in the United States [2]. Importantly, research has consistently demonstrated the direct influence of tobacco POS advertising on tobacco use [1,3]. Exposure to tobacco POS advertising has been positively associated with the urge to smoke and negatively associated with cessation among adult smokers [3]. Exposure to tobacco POS advertising is positively associated with susceptibility, initiation, and current tobacco use among youth and never smokers [4-6]. In communities where there is more tobacco POS advertising, tobacco use is higher; conversely, smoking rates are lower in communities that have adopted policies restricting tobacco POS advertising [7]. Furthermore, the effects of tobacco POS advertising contribute to disparities in smoking and tobacco-related diseases, as potentially vulnerable populations are often explicitly targeted. Key examples include greater tobacco retailer density in communities of color and pervasive POS advertising of menthol cigarettes in lower-income African American communities [8-11].

The regulation of tobacco POS advertising varies substantially across communities and states within the United States [12]. Surveillance of tobacco POS advertising is important for assessing compliance with local, state, and federal regulations; informing evidence-based policy making; understanding industry behavior; and identifying factors contributing to ongoing disparities in tobacco use and tobacco-related disease burden [13]. At present, store audits represent the most common and rigorous approach to measuring tobacco POS advertisements. In general, researchers recommend a store audit to involve the careful observation of advertisements and retail spaces, speaking with store clerks, and manually photographing and annotating the advertisements present in brick-and-mortar tobacco retailers [1,13]. Such audits require substantial resources, training, and time; thus, they may be difficult to conduct in underresourced communities. For example, a 2014 study surveying 48 states found that the majority of surveillance work at the local level was conducted by volunteer staff [14]. Furthermore, at the time of the study, only slightly more than half (54%) of the surveyed states reported conducting surveillance activities in the past several years. Among those conducting surveillance, only 19% reported that these activities were routine, reinforcing the challenges of consistent data collection via traditional surveillance methods and the importance of leveraging new technology and strategies. Technological advancements would thus be beneficial for improving the depth and breadth of in-store surveillance.

Over the past decade, 2 technological innovations have advanced sufficiently to offer the possibility of a more efficient, less resource-intensive approach for measuring POS advertising at scale. Specifically, the combination of crowdsourcing and machine learning offers a promising strategy to automate the store auditing process and allow researchers to gain insights into actual advertisements. Machine learning is a general term to describe a collection of related computer tasks, including image classification and object detection—key elements in identifying tobacco brands in photographs of store environments. One advantage of machine learning is that it eliminates the burden of manual review and coding and carries the potential for major cost and time savings for research projects. Although both traditional surveillance activities and crowdsourcing efforts have been used to collect data from tobacco retailers, machine learning is yet to be applied to improve digital photograph extraction [14-17].

Objectives

The primary aim of this study is to assess the feasibility of using machine learning approaches to identify and quantify information accurately in tobacco POS advertising. The proposed approach has 2 components. First, we make use of a large collection of interior photographs of tobacco retailers in West Virginia and Washington, DC. Second, we apply image classification and object detection to identify the brand, number, and size of tobacco advertisements within digital photographs. With respect to image classification, we aim to identify individual tobacco brands in the archive of photographs. We aim to accurately determine the location of brand-specific advertising images in individual photographs by using object detection. By accurately classifying branded images and automatically detecting the location of branded imagery in a retail environment, our approach can provide a practical alternative to in-person POS store audits. In addition, being able to capture enhanced contextual information about the POS environment may provide insights into efforts to combat ongoing efforts from the tobacco industry to use their advertisements to target vulnerable populations.

Methods

Data Collection

Camera glasses were used by field staff to collect interior photographs from 410 tobacco retailers across 37 counties in West Virginia from September 2013 to August 2014, the majority of which were in 6 counties with full coverage of all stores. A total of 86,683 digital photographs of tobacco product counters and advertisements were collected. These photographs were then used in the first machine learning task, which attempted to classify specific tobacco brands within the photos. We also collected an additional 13,264 photographs from 82 tobacco retailers in Washington, DC, during the fall of 2018 for the second machine learning task, which focused on identifying the location of advertisements within each photograph.

Data Cleaning

We manually sorted the photographs into training and validation sets for purposes of training a neural network to classify the retailer photographs by brand. In so doing, we removed the images that contained no advertising or had unclear information, leaving 0.8% (694/86,683) photographs deemed most useful for brand detection analysis between West Virginia and Washington, DC. Although 0.8% (694/86,683) seems low, our cameras took photographs every 1 second, meaning most did not contain any tobacco POS advertising. Once the 694 clearest photographs were selected, they were manually sorted by which brands were contained within them. Finally, we created a training set of 70% (486/694) of the photographs and a testing/validation set of 30% (208/694) of the photographs for classification.

To detect the location of advertisements within each photograph, only photographs with Marlboro advertisements were ultimately used, but from a larger pool of photographs, as not all brands needed to be clear. We decided only to attempt to detect the location of Marlboro advertisements because they were by far the most common brand, giving us the largest amount of data on which to train our model. A sample of the 843 clearest Marlboro photos across Washington, DC, and West Virginia were sorted into training (589/843, 69.9% photographs) and testing sets (254/843, 30.1% photographs).

Classifying the Presence of Brands in Photographs

Once the images in the training and test sets were manually classified, we used the Python TensorFlow and Keras libraries through the Jupyter Notebooks platform to recreate the manual brand classification with machine learning [18,19]. Fundamentally, our model analyzed each image, resulting in scored and predicted probabilities that a given image contained specific tobacco brands.

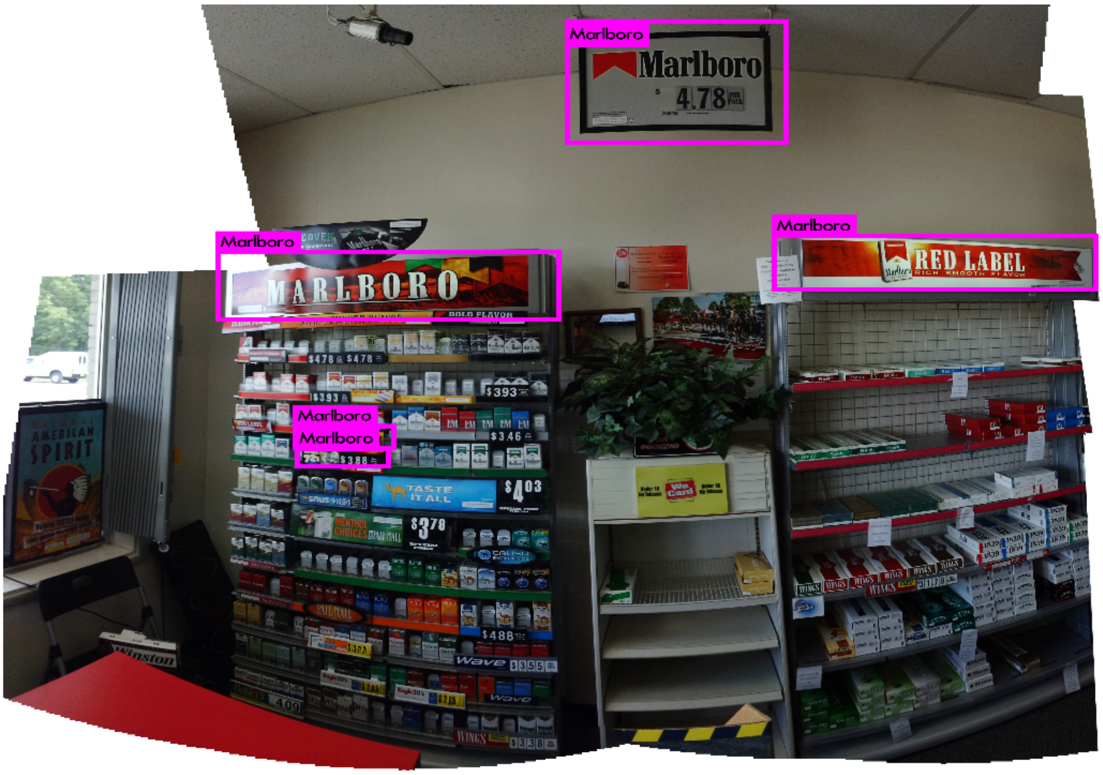
Several steps were taken to speed up learning and minimize the computation time owing to the computationally expensive nature of image classification models. First, all processes were executed on a Linux server hosted by Amazon Web Services. Second, we used a pretrained image classification network, Inception V3, which has already been trained to classify millions of labeled images in the *ImageNet* repository [20-22]. Our brand classification model extends Inception V3 to categorize tobacco brands by training a new classification layer on top of its existing architecture, allowing us to successfully classify brands with considerably fewer images than would otherwise be necessary.

After importing the pretrained Inception V3 classification network, we built a deep learning neural network to classify whether the images contained specific tobacco brands. A deep learning network can recognize features in the images that are associated with a targeted brand, including the colors, shapes, or patterns in a given logo. Specifically, we used TensorFlow to configure our computer graphics processing units before training our neural network using Keras. To prevent overfitting and make the most of the 486 photographs in our training set, we configured several random transformations so that our model would never see the same exact picture twice. We randomly applied zooms, shearing, and horizontal transformations to our training set and processed 50 images per batch at a resolution of 299×299 pixels. Once the neural network was trained, we generated the predicted probabilities that each image contained a brand of interest. In our analysis, an image was classified as containing the logo of the brand if the probability was ≥0.5, but other cutoffs could have been chosen.

Discovering Object Location Within Images

Beyond identifying the presence of specific brands, our second goal is to discover their locations in a given photograph to inform their positioning and density in an individual store. Doing so required a different technology than the Inception V3 classification network described earlier, which was designed to discover logos anywhere in an image. We trained YOLO (You Only Look Once) V3, a state-of-the-art, real-time unified object detection system to detect tobacco advertisements within images, as illustrated in Figure 1 [23]. Compared with region-proposal-based convolutional neural networks (eg, R-CNN [region-based convolutional neural networks], fast R-CNN, and faster R-CNN), YOLO V3 uses a single convolutional neural network optimized end to end to full images to simultaneously predict multiple bounding boxes and their class probabilities. By being informed of the global context of the image, YOLO V3 has shown the ability to predict fewer false positives in background image areas where objects are not present [24].

Figure 1. Example image from a tobacco point of sale with YOLO (You Only Look Once) bounding boxes.



As part of our process, we first used the open-source Visual Object Tagging Tool to draw bounding boxes around each Marlboro advertisement within images and generate related annotations for training [25]. We then trained YOLO V3 using a pretrained, publicly available Darknet-53 model on ImageNet to perform customized object detection [24,26]. Anchor boxes, predefined boxes to improve speed and efficiency in detection of typical objects of interest, were recomputed using K-means clustering with intersection over union (IOU) as the distance measure. We stopped training at the 10,000th batch, where the training loss levels off. To reduce overfitting and generalization error, we tested the network using weights from alternative stopping points generated earlier via the testing data set.

We then evaluated the model accuracy for location detection using testing images based on two metrics: the mean average precision (mAP) and IOU. The mAP is a composite accuracy indicator that ranges from 0 to 1 and accounts for both precision and recall, which is computed as the area under the precision-recall curve. An mAP score of 1 indicates that 100% of the model's predictions are correct and that 100% of the truth objects are detected by the model. The IOU measures the extent to which the prediction overlaps with the ground truth, which is given by the ratio of the area of intersection and area of union of the predicted bounding box and ground-truth bounding box.

Results

Overview

We present our results following our 2 primary research questions, as outlined earlier. First, we describe the success of identifying individual tobacco brands in our set of photographs. Second, we detail how we determined the location of such images in individual stores. We then discuss our findings and their implications in the *Discussion* and *Conclusions* sections.

Brand Detection

Our Inception V3 model fundamentally generated the predicted probabilities of the presence of each brand for each image in the validation set. Although we ultimately decided to focus on the brand Marlboro, we attempted to detect a series of brand logos. Success varied by brand, but our model achieved a classification accuracy of more than 75% for 8 of the 42 brands it was trained to detect (Textbox 1 and Table 1). For all but 7 brands—Camel, Marlboro, Pyramid, Pall Mall, Grizzly, Swisher, and Newport—the number of labeled example images constituted less than 11.7% (57/486) of the training data set. Table 2 shows the predicted probabilities of discovering specific logos in Figures 2 and 3 to illustrate the variability in the size and design of advertisements and the ability of the model to interpret branding in complex images. As shown, Newport was predicted with the highest probability in Figure 2, with Marlboro having the highest probability in Figure 3, with other brands still having high probabilities in each. Each image varies in terms of the heterogeneity and scale of the logos in question, implying the importance of both color and design.

Textbox 1. Tobacco brands considered for classification.

Cigarette

- Marlboro
- Newport
- Camel
- Pall Mall
- Pyramid
- Maverick
- Santa Fe
- Winston
- Kool
- American Spirit

Cigar

- Swisher
- Black and Mild
- White Owl
- Dutch Masters
- Winchester
- Garcia y Vega
- Phillies
- Cheyenne
- Backwoods

Smokeless tobacco

- Levi Garrett Plug
- Day’s Work
- Red Man Plug
- Grizzly
- Garrett
- Skoal
- Red Man
- Copenhagen
- Red Seal
- Timberwolf
- Kayak
- Beechnut
- Kodiak
- Longhorn

Snus

- Skoal
- General

e-Cigarettes

- Blu

- FIN
- Logic
- MARKTEN
- NJOY
- V2
- VUSE

Table 1. Classification accuracy for validation data by tobacco brands.

Brand	Classification accuracy (%)
Copenhagen	90.4
Winston	85.1
Pyramid	82.7
Blu	81.7
American Spirit	80.3
Marlboro	78.8
Camel	75.9
Pall Mall	75.9

Table 2. Predicted probability of logos by Inception V3.

Photograph and brand	Probability
Figure 2	
Marlboro	0.635
Newport	0.982
Camel	0.661
Figure 3	
Marlboro	0.993
Newport	0.868

Figure 2. Panoramic image of a tobacco point of sale—view 1.

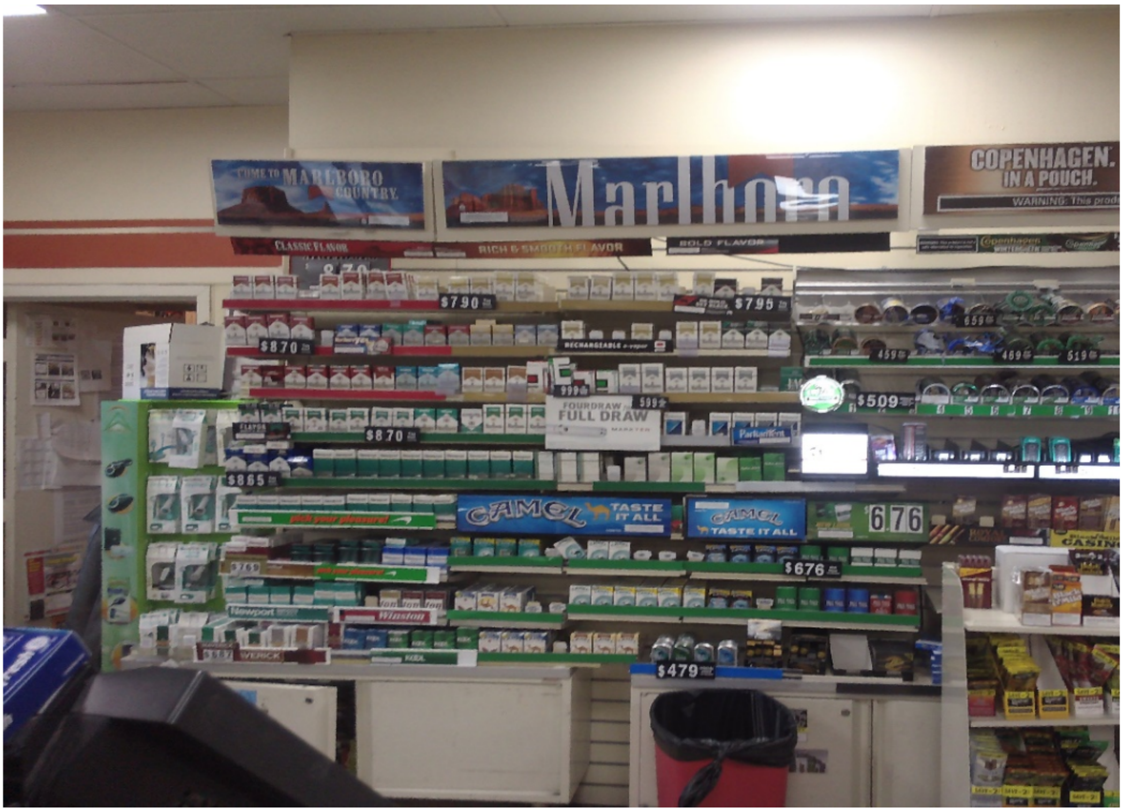
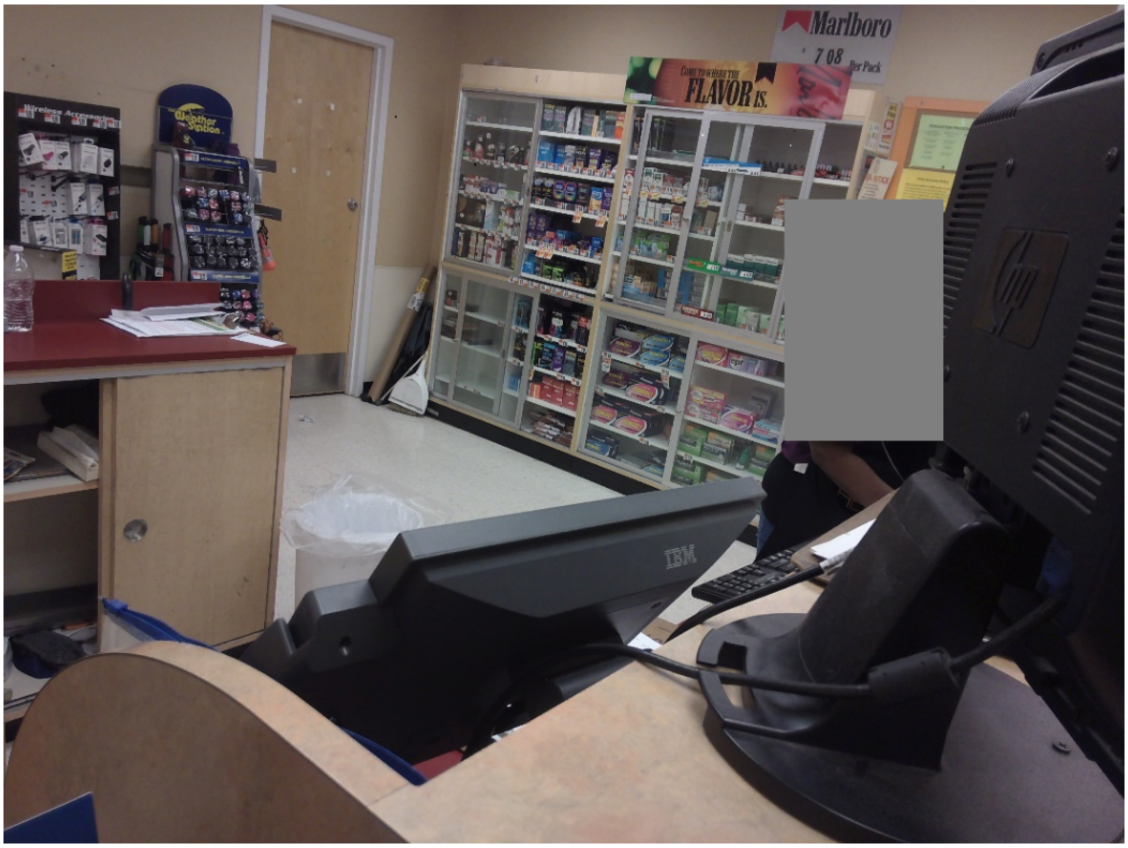

COME TO MARLBORO COUNTRY
Marlboro
COPENHAGEN. IN A POUCH.
CLASSIC FLAVOR
RICH & SMOOTH FLAVOR
BOLD FLAVOR
$790
$795
$870
$865
$509
FULL DRAW
CAMEL TASTE IT ALL
6.76
$676
Newport
Winston
$479

Figure 3. Panoramic image of a tobacco point of sale—view 2.


Marlboro
7.08 Per Pack
FLAVOR IS.
IBM

[redacted]

Location of Advertisement Detection

As described in the *Methods* section, for the purpose of location detection, we evaluated YOLO V3 model accuracy using the mAP and IOU on testing images. The network with weights that yielded the highest testing mAP (0.72; Figure 4) and IOU score (0.62; Figure 4) was chosen as the best model for detection. Figure 5 shows the detection results for an example image, where five Marlboro advertisements were detected. The object detector also generated a confidence score for each box, along with estimates of the upper left coordinates and the absolute width and height of the bounding boxes (Table 3).

Figure 4. Training loss and testing accuracy of the YOLO (You Only Look Once) V3 objection detector. IOU: intersection over union; mAP: mean average precision.

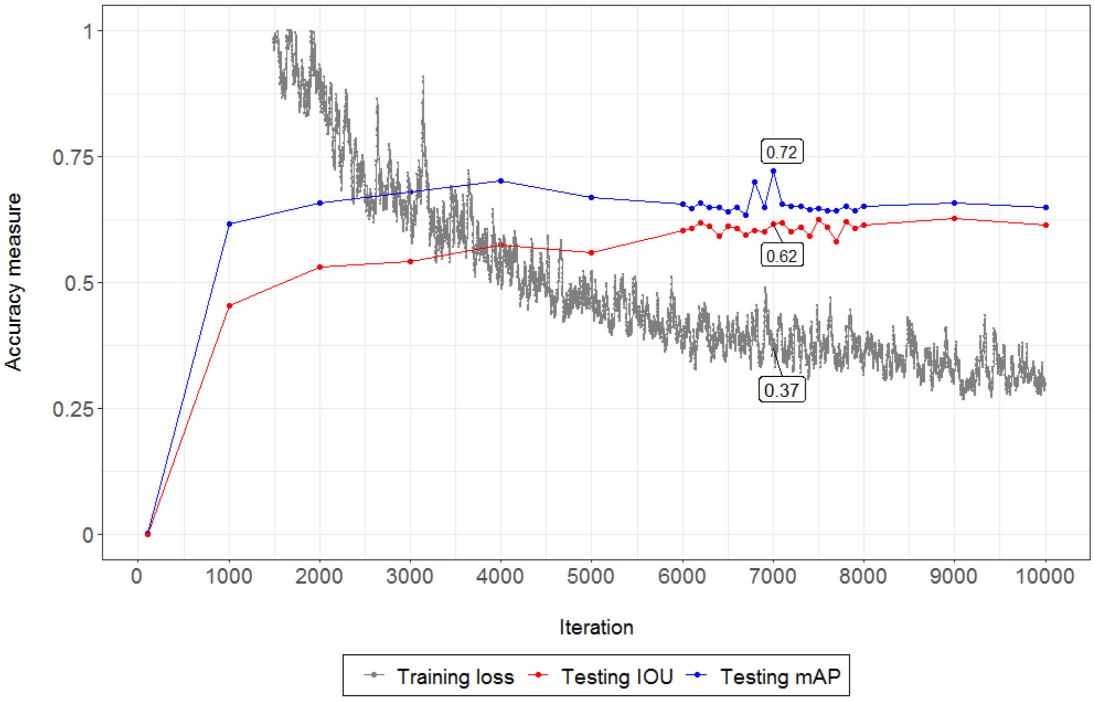


Figure 5. Prediction of Marlboro signs by YOLO (You Only Look Once) V3.

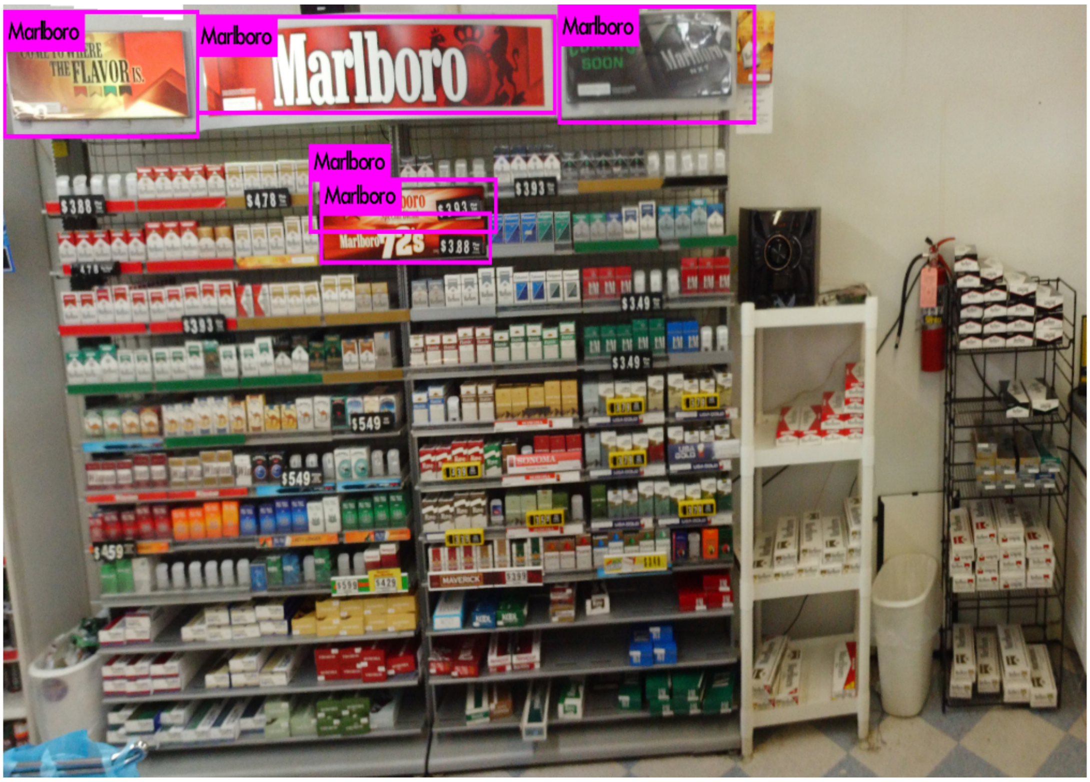


Table 3. Bounding boxes of Marlboro signs detected by YOLO (You Only Look Once) V3.

Bounding box in Figure 5	Confidence score (%)	Bounding box measures (pixels)			
		Left x	Top y	Width	Height
1	100	0	16	737	376
2	100	724	30	1355	293
3	98	1149	505	708	163
4	100	1189	604	648	154
5	100	2084	0	745	358

Discussion

Principal Findings

Our study provides evidence of the feasibility of a novel methodological approach that tobacco researchers and other public health practitioners can apply in the collection and processing of data for tobacco and other POS surveillance efforts. Trained on a small set of labeled tobacco POS photographs, our classifier and object detector were able to identify tobacco brands and their location and dimension successfully within images. Although the initial labeling of the training data set was time-consuming, with an average of 3 minutes per photograph for staff to label advertisements, the costs of processing the photographic data decreased once the neural networks were trained. We were able to achieve a brand classification accuracy of over 75% for 8 of the 42 brands that our classifier (Inception V3) was trained to detect. Furthermore, we were able to accurately predict the location of branded advertising within the store environment (YOLO V3). Our location predictions achieved an mAP score of 0.72 and an IOU score of 0.62.

Our Inception V3 brand detection model had 2 major characteristics. First, the model was optimized to predict true positives, but with the unintended consequence of predicting more false positives. The model, therefore, tended to decide that a brand was in a photograph when none was present, in the process of finding true positives. Unfortunately, there was no alternative; tuning the model to favor predicting true negatives would have been too conservative (ie, unable to make any prediction on most photographs). Second, the model was heavily dependent on making predictions based on color. For example, in Figure 2, the size of the Marlboro advertisement is much larger than that in Figure 3. However, the predicted probability of a Marlboro advertisement in Figure 3 is higher owing to the presence of a familiar red logo, as shown in Table 1. The presence of green on the shelves of the menthol product displays may have caused the very high predicted probability of Newport in both photos, despite the Newport advertisements themselves being relatively small. Such variation in advertisement design and color, especially for Marlboro, may also explain why the accuracy of predicting the presence of Marlboro advertisements was among the lowest, despite having the most representation across photographs. However, our results indicate the potential for accurately identifying the presence of specific tobacco brands in digital photographs using Inception V3 or similar models. Although the accuracy rates shown above do not approach some in the medical field with very large training data sets, they do show how our approach is considerably more effective than random initialization constraints, even with a relatively small training set [27]. For context, a data set containing at least 2000 images of each brand with varying sizes, rotation angles, lighting schemes, and backgrounds would be needed for improved object detection accuracy [28].

This promising technology offers potential opportunities for tobacco POS surveillance to move forward. First, conducting timely, long-term surveillance of the POS environment can be resource intensive. Some existing state and local audit systems require a substantial amount of training and resources; however, they do provide critical information for enforcement activities as well as an understanding of the impact of marketing on current tobacco user behaviors and tobacco use initiation [1,3,13]. By applying machine learning techniques for efficient image classification and object detection, the methods described in this paper can assist in ensuring that retailers comply with specific retail provisions, such as state or local flavored tobacco bans, in addition to potentially improving the generalizability of research results by increasing the reliability and standardization of advertising information and classification.

The use of improved surveillance technologies can also assist in the assessment of local policy impacts in an evolving POS retail environment. With the adoption of local and state restrictions on the sale of flavored tobacco products, jurisdictions have an ever-growing need to assess the impact of these policies on tobacco retail environments [13]. For example, more routine and detailed retail POS advertising data could be used to assess differences before and after policies are implemented to examine key questions regarding changes in product availability, advertising, and promotion or discounting efforts.

Finally, this concept of applying machine learning to examine tobacco POS could also be extended to other public health applications. For example, image classification and object detection can be applied to identify other products of interest such as sugary drinks, candy, and processed foods from retail store images. Such an approach would allow public health researchers and policy makers to gauge the prevalence and types of advertisements for each product and understand how a specific retail food environment may interact with population demographics. Given the current attention on obesity, related health outcomes, and efforts to tax sugary drinks, we might anticipate interest in a machine learning–based approach, as illustrated [29].

Limitations

Despite the success of applying deep neural networks to tobacco advertisement object detection, our analysis revealed several challenges, especially with respect to object location detection. Although our brand classification algorithms were generally successful (Inception V3) with some false-positive classifications, object location detection (YOLO V3) was more challenging, requiring a large amount of training data to identify patterns. Owing to the relatively small sample size, we decided to limit training our object detection algorithm to Marlboro advertisements, the dominant tobacco brand in our data set. Even when considering one brand within a given POS, we observe substantial variability in advertising content and appearance, including differences in color, shading, perspective, size, and rotation (Figure 5). Such variability made it difficult for object detection algorithms to learn and recognize brand patterns. Therefore, object location detection can be considered a more complex task prone to false negatives (ie, missing an object that is actually present). Our object detection model is conservative in its present form, with the side effect of missing true objects present in a photo, in contrast to our brand recognition model that generated false positives. Either bias would be improved by training the image classifier or object detector with additional images but would require further resources.

One key factor that limited our training data set is that relatively few tobacco POS advertising training images are available to the public, as distinct from generic computer vision applications where large and clean benchmark data sets exist. Collecting primary data on tobacco advertising and creating sufficient labeled training samples are a challenge in performing deep neural network–based object detection owing to time and cost implications. As shown in the literature, primary data collection for POS surveillance poses many challenges related to technology, workforce training, and overall logistics [16]. Data availability presents a particular concern for detecting tobacco brands that are less common in the marketplace, such as vape products, which are of increasing concern to researchers and health officials. Owing to the rapid evolution of branding, marketing, and delivery of tobacco products, we expect such analytical and data collection challenges to persist. Such challenges are further complicated by the lack of a centralized data store to track advertising materials. No comprehensive, publicly available, federal repository for tobacco advertising materials currently exists, although efforts by academic and nonprofit institutions to conduct marketing surveillance persist, including the Rutgers Center for Tobacco Studies and Campaign for Tobacco-Free Kids [30,31].

Beyond technology, there are also specific practical considerations when implementing such assessments. First, there are privacy and safety concerns associated with the use of camera glasses to collect POS photographs in public, especially when one or more identifying people are incidentally captured in photographs. Many store owners may not be comfortable with individuals using camera phones or handheld cameras to capture POS information. To accommodate this, camera glasses are often used to capture hands-free images in a more inconspicuous way. However, wearing a hidden camera in public may raise legal issues, depending on the location. For example, although capturing public scenes requires no consent in the United States, the opposite exists in Spain [32]. Although consent is not required in the United States for public photography, it may not always be socially acceptable. Clear guidelines about the “dos and don’ts” of wearing camera glasses for data collection should be created to broaden the applications of such technology to POS surveillance.

Finally, data quality must be considered when collecting data for machine learning and related analysis. As we discovered, cameras in glasses may not always capture high-quality or usable images. Photographs may be blurry, overexposed, or missing the desired images. As such, input data quality may be subject to bias and error, which could affect the development of subsequent machine learning algorithms. In addition, because many glasses do not have a remote trigger or a viewing screen, the individual using the glasses may be unaware of the image quality until the photos are uploaded. Such effects could result in a significant reduction in the availability of the training images used for image classification and object detection. However, having multiple overlapping images of any given POS location helps to ensure that all available POS advertising is captured with one or more clear photographs.

Future work on image classification in the tobacco POS retail environment would benefit from exploring additional methods to reduce false positives and false negatives for respective machine learning algorithms. For example, we could take advantage of image repositories that contain tobacco branding to help train models and add additional nuances. Although such repositories would not provide sufficient data to help describe the US retail environment comprehensively, they could assist by providing different geographical contexts or supplemental material in other advertising channels (eg, magazines, newspapers, films, the web, and social media) may still be helpful [33,34]. The introduction of crowdsourced data collection for use with machine learning techniques, as well as incorporating social media data, may also assist in gaining new sources of data and reducing the logistical burdens of surveillance programs. Integrating these efforts into a shared repository of tobacco-related images among tobacco control programs would also assist in improving brand identification and location detection and training of new algorithms. Such collaboration among local, state, federal, and academic institutions could be a powerful tool in understanding fast-changing retail trends and regional heterogeneity. In addition, our study considered location to be relative to each photograph, rather than in a defined geographic location within the POS. Although tobacco advertising tends to be consistent in the retail environment, such as in the commonly used Power Wall, value could be added by considering the purely geographic factor in future analyses [35].

Conclusions

To summarize, our study demonstrated the utility of machine learning for POS assessment and highlighted some existing technological limitations. Although the current machine learning algorithms are advanced, they still have room for improvement. Large-scale POS photographic data sets that are comprehensive enough to capture a wider range of tobacco brands are needed to train multiclass algorithms capable of detecting advertisements from less common tobacco brands. Future studies should also attempt to detect the prices associated with within-store advertisements by incorporating text recognition algorithms to gain more contextual information on tobacco marketing. Attempting to classify an absolute geographic location within stores instead of relative image location within a photograph may also be of interest to researchers and surveillance efforts.

Despite these limitations, our analysis shows the promise of using machine learning as part of a suite of tools to understand the tobacco retail environment and inform public health interventions at multiple scales. The accurate and automatic classification of product brands and detection of their location within a retail environment could assist in developing a practical alternative to in-person POS audits, especially in resource-limited environments. For example, the necessary classifiers—documentation—could be made available in the public domain to facilitate their use by public health departments. In addition, coded photographs could be shared as part of a centralized resource to reduce the level of effort required to conduct or continue similar evaluations. With the increasing sales restrictions at the POS, surveillance products with enhanced contextual information about the retail environment can provide states, counties, and municipalities the opportunity to better understand the impact of existing and proposed policies, including ongoing efforts by the tobacco industry to target potentially vulnerable populations.

Acknowledgments

This work was supported in part by Truth Initiative and included activities such as data collection, analysis, interpretation of data, and writing the manuscript. In addition, any store photos located in Washington, DC, that were used in this research were collected from a study supported by the National Cancer Institute of the National Institutes of Health under award number R21CA208206. Although AAR is a United States Food and Drug Administration's Center for Tobacco Products employee, this work was not done as part of his official duties. This publication reflects the views of the author and should not be construed as reflecting the United States Food and Drug Administration's Center for Tobacco Products' views or policies.

Conflicts of Interest

None declared.

References

1. 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]
2. Federal Trade Commission Cigarette Report for 2018. Washington, DC: FTC; 2019. URL: https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-cigarette-report-2018-smokeless-tobacco-report-2018/p114508cigarettereport2018.pdf
3. Wakefield M, Germain D, Henriksen L. The effect of retail cigarette pack displays on impulse purchase. Addiction 2008 Feb;103(2):322-328. [doi: 10.1111/j.1360-0443.2007.02062.x] [Medline: 18042190]
4. Robertson L, Cameron C, McGee R, Marsh L, Hoek J. Point-of-sale tobacco promotion and youth smoking: a meta-analysis. Tob Control 2016 Dec;25(e2):e83-e89. [doi: 10.1136/tobaccocontrol-2015-052586] [Medline: 26728139]
5. Henriksen L, Feighery EC, Schleicher NC, Cowling DW, Kline RS, Fortmann SP. Is adolescent smoking related to the density and proximity of tobacco outlets and retail cigarette advertising near schools? Prev Med 2008 Aug;47(2):210-214. [doi: 10.1016/j.ypmed.2008.04.008] [Medline: 18544462]
6. 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]
7. 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]
8. Widome R, Brock B, Noble P, Forster JL. The relationship of neighborhood demographic characteristics to point-of-sale tobacco advertising and marketing. Ethn Health 2013;18(2):136-151. [doi: 10.1080/13557858.2012.701273] [Medline: 22789035]
9. Loomis BR, Kim AE, Busey AH, Farrelly MC, Willett JG, Juster HR. The density of tobacco retailers and its association with attitudes toward smoking, exposure to point-of-sale tobacco advertising, cigarette purchasing, and smoking among New York youth. Prev Med 2012 Nov;55(5):468-474. [doi: 10.1016/j.ypmed.2012.08.014] [Medline: 22960255]
10. Lee JG, Henriksen L, Rose SW, Moreland-Russell S, Ribisl KM. A systematic review of neighborhood disparities in point-of-sale tobacco marketing. Am J Public Health 2015 Sep;105(9):e8-18 [FREE Full text] [doi: 10.2105/AJPH.2015.302777] [Medline: 26180986]
11. Ribisl KM, D'Angelo H, Feld AL, Schleicher NC, Golden SD, Luke DA, et al. Disparities in tobacco marketing and product availability at the point of sale: Results of a national study. Prev Med 2017 Dec;105:381-388 [FREE Full text] [doi: 10.1016/j.ypmed.2017.04.010] [Medline: 28392252]

12. Point-of-Sale Strategies: A Tobacco Control Guide. St. Louis, MO: George Warren Brown School of Social Work at Washington University in St. Louis and the Tobacco Control Legal Consortium; 2014.
13. Lee JGL, Henriksen L, Myers AE, Dauphinee AL, Ribisl KM. A systematic review of store audit methods for assessing tobacco marketing and products at the point of sale. Tob Control 2014 Mar;23(2):98-106 [FREE Full text] [doi: 10.1136/tobaccocontrol-2012-050807] [Medline: 23322313]
14. Point-of-Sale Report to the Nation: Policy Activity 2012-2014. St. Louis, MO: Center for Public Health Systems Science at the Brown School at Washington University in St. Louis and the National Cancer Institute, State and Community Tobacco Control Research Initiative; 2015.
15. Mortensen K, Hughes TL. Comparing amazon's mechanical Turk platform to conventional data collection methods in the health and medical research literature. J Gen Intern Med 2018 Apr;33(4):533-538 [FREE Full text] [doi: 10.1007/s11606-017-4246-0] [Medline: 29302882]
16. Cantrell J, Ganz O, Ilakkuvan V, Tacelosky M, Kreslake J, Moon-Howard J, et al. Implementation of a multimodal mobile system for point-of-sale surveillance: lessons learned from case studies in Washington, DC, and New York city. JMIR Public Health Surveill 2015;1(2):e20 [FREE Full text] [doi: 10.2196/publichealth.4191] [Medline: 27227138]
17. Ilakkuvan V, Tacelosky M, Ivey KC, Pearson JL, Cantrell J, Vallone DM, et al. Cameras for public health surveillance: a methods protocol for crowdsourced annotation of point-of-sale photographs. JMIR Res Protoc 2014;3(2):e22 [FREE Full text] [doi: 10.2196/resprot.3277] [Medline: 24717168]
18. Chollet F. Keras. Github Repository. 2017. URL: https://github.com/fchollet/keras/ [accessed 2020-09-12]
19. Kluyver T, Ragan-Kelley B, Pérez F, Granger B, Bussonnier M, Frederic J, et al. Jupyter Notebooks—a Publishing Format for Reproducible Computational Workflows. In: Proceedings of the 20th International Conference on Electronic Publishing. 2016 Presented at: CEP'16; June 17-19, 2016; Amsterdam. [doi: 10.3233/978-1-61499-649-1-87]
20. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the Inception Architecture for Computer Vision. In: Proceedings of the IEEE conference on computer visionpattern recognition (CVPR). 2016 Presented at: IEEE'16; June 27-30, 2016; Las Vegas, Nevada URL: http://doi.ieeecomputersociety.org/10.1109/CVPR.2016.308 [doi: 10.1109/cvpr.2016.308]
21. Donahue J, Jia Y, Vinyals O, Hoffman J, Zhang N, Tzeng E, et al. DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition. In: Proceedings of the 31st International Conference on Machine Learning. 2014 Presented at: ML'14; June 22-24 2014; Beijing p. 647-655.
22. Summary and Statistics. Image Net. 2010. URL: https://image-net.org/about.php [accessed 2020-09-11]
23. Redmon J, Divvala S, Girshick R, Farhadi A. You Only Look Once: Unified, Real-time Object Detection. arXiv. 2016. URL: https://arxiv.org/abs/1506.02640 [accessed 2020-09-10]
24. Redmon J, Farhadi A. Yolov3: An Incremental Improvement. arXiv. 2018. URL: https://arxiv.org/abs/1804.02767 [accessed 2020-09-10]
25. Microsoft/VoTT. GitHub Repository. 2018. URL: https://github.com/microsoft/VoTT [accessed 2020-09-10]
26. Redmon J. Darknet: Open Source Neural Networks in C. PJ Reddie. 2013. URL: https://pjreddie.com/darknet/ [accessed 2020-09-10]
27. Kermany DS, Goldbaum M, Cai W, Valentim CC, Liang H, Baxter SL, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell 2018 Feb 22;172(5):1122-31.e9. [doi: 10.1016/j.cell.2018.02.010] [Medline: 29474911]
28. Bochkovskiy A. Darknet. Github Repository. 2020. URL: https://github.com/AlexeyAB/darknet [accessed 2020-09-09]
29. Malik VS, Popkin BM, Bray GA, Despre´s J, Hu FB. Sugar-sweetened beverages, obesity, type 2 diabetes mellitus, and cardiovascular disease risk. Circulation 2010 Mar 23;121(11):1356-1364. [doi: 10.1161/circulationaha.109.876185]
30. Center for Tobacco Studies. Rutgers University School of Public Health. URL: http://sph.rutgers.edu/centers/cts/index.html [accessed 2020-09-09]
31. Campaign for Tobacco-Free Kids. URL: https://www.tobaccofreekids.org/ [accessed 2021-01-12]
32. Country Specific Consent Requirements. Wikimedia Commons. URL: https://commons.wikimedia.org/wiki/Commons:Country_specific_consent_requirements [accessed 2020-09-16]
33. Stanford University Research into The Impact of Tobacco Advertising. URL: http://tobacco.stanford.edu/tobacco_main/index.php [accessed 2021-01-11]
34. Smoke Free Movies. URL: https://smokefreemovies.ucsf.edu/ [accessed 2021-01-11]
35. Martino SC, Setodji CM, Dunbar MS, Shadel WG. Increased attention to the tobacco power wall predicts increased smoking risk among adolescents. Addict Behav 2019;88:1-5 [FREE Full text] [doi: 10.1016/j.addbeh.2018.07.024] [Medline: 30098502]

Abbreviations

IOU: intersection over union
mAP: mean average precision
POS: point-of-sale

R-CNN: region-based convolutional neural networks
YOLO: You Only Look Once

Edited by R Kukafka; submitted 18.09.20; peer-reviewed by B Heckmann, P Banik; comments to author 21.12.20; revised version received 18.02.21; accepted 21.06.21; published 27.08.21.

Please cite as:
English N, Anesetti-Rothermel A, Zhao C, Latterner A, Benson AF, Herman P, Emery S, Schneider J, Rose SW, Patel M, Schillo BA
Image Processing for Public Health Surveillance of Tobacco Point-of-Sale Advertising: Machine Learning–Based Methodology
J Med Internet Res 2021;23(8):e24408
URL: https://www.jmir.org/2021/8/e24408
doi:10.2196/24408
PMID:34448700

©Ned English, Andrew Anesetti-Rothermel, Chang Zhao, Andrew Latterner, Adam F Benson, Peter Herman, Sherry Emery, Jordan Schneider, Shyanika W Rose, Minal Patel, Barbara A Schillo. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Predicting Kidney Graft Survival Using Machine Learning Methods: Prediction Model Development and Feature Significance Analysis Study

Syed Asil Ali Naqvi[1], MCS; Karthik Tennankore[2], MD; Amanda Vinson[2], MD; Patrice C Roy[1], PhD; Syed Sibte Raza Abidi[1], PhD

[1]Department of Computer Science, Dalhousie University, Halifax, NS, Canada
[2]Division of Nephrology, Dalhousie University, Halifax, NS, Canada

Corresponding Author:
Syed Asil Ali Naqvi, MCS
Department of Computer Science
Dalhousie University
Goldberg Computer Science Bldg
6050 University Ave
Halifax, NS, B3H 1W5
Canada
Phone: 1 9023290504
Email: a.naqvi@dal.ca

Abstract

Background: Kidney transplantation is the optimal treatment for patients with end-stage renal disease. Short- and long-term kidney graft survival is influenced by a number of donor and recipient factors. Predicting the success of kidney transplantation is important for optimizing kidney allocation.

Objective: The aim of this study was to predict the risk of kidney graft failure across three temporal cohorts (within 1 year, within 5 years, and after 5 years following a transplant) based on donor and recipient characteristics. We analyzed a large data set comprising over 50,000 kidney transplants covering an approximate 20-year period.

Methods: We applied machine learning–based classification algorithms to develop prediction models for the risk of graft failure for three different temporal cohorts. Deep learning–based autoencoders were applied for data dimensionality reduction, which improved the prediction performance. The influence of features on graft survival for each cohort was studied by investigating a new nonoverlapping patient stratification approach.

Results: Our models predicted graft survival with area under the curve scores of 82% within 1 year, 69% within 5 years, and 81% within 17 years. The feature importance analysis elucidated the varying influence of clinical features on graft survival across the three different temporal cohorts.

Conclusions: In this study, we applied machine learning to develop risk prediction models for graft failure that demonstrated a high level of prediction performance. Acknowledging that these models performed better than those reported in the literature for existing risk prediction tools, future studies will focus on how best to incorporate these prediction models into clinical care algorithms to optimize the long-term health of kidney recipients.

(J Med Internet Res 2021;23(8):e26843) doi:10.2196/26843

KEYWORDS

kidney transplantation; machine learning; predictive modeling; survival prediction; dimensionality reduction; feature sensitivity analysis

Introduction

Background

Kidneys are vital for the health of an individual, as they filter waste products from the blood and produce hormones and urine [1]. Patients are considered to have end-stage renal disease when their kidney function falls below a specific threshold [2]. A lack of timely measures to prevent kidney failure results in premature death [3,4].

Kidney transplantation [5,6] and dialysis are the two main treatments for kidney failure [7]. Kidney transplantation offers a survival advantage compared with other forms of kidney replacement therapy; however, the rate of graft loss following transplant is still undesirably high [8]. Kidneys are a limited resource, and optimizing the match between donors and recipients is crucial for improving outcomes after transplantation. Kidney transplant allocation is, in part, based on a number of donor-recipient–related factors that influence graft survival. Various clinical studies have been conducted on the influence of these factors on graft survival; however, given the complex interactions between these factors, there remains much to be learned in this area. Existing risk prediction models only have a limited ability to predict outcomes for kidney transplant recipients with receiver operating characteristic scores of 0.6-0.7 [9-11].

Prediction modeling using machine learning (ML) algorithms has gained attention in recent years [12] for predicting the success of clinical or surgical procedures (such as kidney transplant). ML algorithms autonomously learn the underlying associations between preprocedure clinical features and postprocedural outcomes to predict the outcome of the procedure for a given clinical case. In kidney transplant, ML-based prediction models, based on donor-recipient information, autonomously learn the underlying relationships between donor and recipient factors to predict transplant outcomes. Multiple studies have been conducted using ML methods to predict the kidney graft outcome [13-16]; however, the standard approach in nearly all the reviewed studies has been to select one or more arbitrary period starting from the date of transplant and applying classification algorithms for prediction. There is a clear need for further exploration of data stratification approaches and other ML methods with respect to feature engineering and prediction modeling.

Objectives

In this study, the intent is to investigate kidney transplant allograft survival, that is, estimating the time-to-event and the evolving influence of clinical features leading to an event—within three temporal cohorts of 1 year, >1-5 years, and >5 years of a kidney transplant. We predicted the outcome of graft failure after kidney transplant based on the analysis of donor and recipient features. We applied ML methods to (1) predict the graft status over different temporal periods and (2) analyze the changing effect of donor-recipient–related predictors across different periods. To develop the prediction models, we analyzed a large data set of over 50,000 transplants covering approximately a 20-year period of kidney transplants in the United States. To generate the clinically meaningful temporal cohorts, we experimented with two patient stratification approaches: (1) a novel *nonoverlapping* patient stratification approach, whereby a patient's graft failure was recorded only in the temporal cohort when it actually happened, that is, a graft failure event in the preceding cohort was not included and (2) the traditional *overlapping* patient stratification approach that provides an accumulative count of graft outcomes until a specific time point. To develop the prediction models, we investigated multiple ML algorithms using both patient stratification approaches. Nonoverlapping temporal cohorts were considered to investigate the influence of clinical features on predicting graft survival over time, as the temporal partitioning of the data allowed for the establishment of feature influence across distinct temporal windows. We applied the feature importance method based on the mean decrease in impurity (Gini).

The contributions of this research are as follows: (1) ML-based prediction models that are trained on a large data set, offering improved prediction performance compared with previous studies (previous graft prediction studies are based on a smaller number of transplants over a shorter period); (2) data dimensionality reduction based on a deep learning framework to handle the high-dimensional and complex kidney transplant data set; (3) a novel nonoverlapping patient stratification approach to provide fine-grained feature importance within a specific period while avoiding bias from preceding cohorts; (4) explaining the influence of the different clinical features, during different periods, toward the prediction performance of ML prediction models. This finding allows the selection of the most important features to predict graft outcomes within a specific temporal window; and (5) a comparison between the two stratification approaches with respect to the performance of the prediction models. The future practical outcome of this study is the provision of a data-driven decision support tool to assist nephrologists in the kidney allocation process by identifying the best donor and recipient pair that will lead to the highest likelihood of graft survival for a given recipient.

Prior Work

Patients can receive a kidney from either deceased donors or living donors. The donor-recipient matching process becomes relatively more complex with deceased donors because of the need to account for additional clinical factors (ie, prolonged cold ischemia time, prolonged wait times, and generally lower quality organs) [17]. Given the fact that kidney organs are a limited resource, it is important to have an efficient and effective donor-recipient matching process to ensure long-term graft survival [18].

Data-driven methods are now being used for organ matching; these methods are used to establish clinical compatibility beyond the blood group and tissue type. Conventional data-driven prediction methods use statistical techniques such as Cox proportional hazard models and Kaplan-Meier estimates to perform time-to-event analyses [19]. Significant research has been conducted with Cox-based models in the survival analysis of different organ transplants; however, these methods eventually lose predictive accuracy as the feature space increases [13,14].

ML-based data analysis to develop prediction models for predicting outcomes is usually performed using classification methods, whereas regression methods are used for time-to-event analysis. There are two prominent approaches to predict kidney allograft outcomes using ML-based classification methods. The first approach is to predict graft survival over time by dividing a longitudinal data set into different time cohorts based on the occurrence of a given adverse event or the last follow-up date from the date of transplant. Each time cohort has a binary target variable, that is, success or failure of the graft, which is used to train the classification model to predict graft survival [15,16]. The second approach is to predict the risk acuity associated with a graft within a period. Topuz et al [15] used this approach and divided the data set into three graft failure risk groups (high, medium, and low) across three different periods to predict the risk of graft failure within a specific time using classification methods. Li et al [20] used Bayes net to classify graft risk levels and predict graft survival.

Due to the high dimensionality of existing data sets for organ transplantation, feature selection is applied to filter out redundant features. A stacked autoencoder, which is an unsupervised neural network, is an efficient dimensionality reduction technique with promising performance for deep representation of medical data [21] that reconstructs its own inputs by first encoding them to a smaller size and then decoding back to the original inputs. A comparative study by Sadati et al [22] highlighted the efficacy of different types of autoencoders for data sets based on electronic health records.

Right-censored data are a common problem for survival analysis, as it represents cases for which the adverse event is not available or recorded because of either the subject having been lost to follow-up or not experiencing the event during the study period. Multiple approaches have been adopted in previous studies to address this problem. The study on kidney transplants by Topuz et al [15] discarded all the right-censored data before 7 years from the time of transplant and included the remaining transplants that took place after that time point in the low-risk group. In a study predicting heart transplant outcomes, the data set was divided into three different time cohorts (1, 5, and 9 years) to predict the status of the graft. Patients who did not have graft failure during a particular time cohort were censored, and all the patients beyond that time cohort were considered to have successful transplants [23].

The influence of clinical features (or clinical predictors) on graft survival tends to vary over time [16], as shown using a heat map [24]. Dag et al [23] analyzed the changing significance of features for three overlapping time cohorts (1, 5, and 9 years). They deduced that certain types of features perform well in the long term compared with the short and medium terms. For instance, socioeconomic factors were more influential in their 9-year time cohort as they covered major variations in the data. It is important to note here that feature significance over time can only be substantiated if the analysis is performed with nonoverlapping cohorts to avoid any bias introduced by the cumulative effect of data before the analysis period.

In previous studies [16,20-22,25], predicting graft failure has been pursued by taking an overlapping patient stratification approach, which means that each subsequent time cohort includes data from the previous cohort. This introduces a cumulative effect that is useful for predicting graft failure across a staggered time period. However, the overlapping patient stratification approach is ineffective in determining the influence of clinical features during a specific time period. Hence, the nonoverlapped cohort approach offers a novel mechanism to investigate the influence of clinical features within specific time windows. To the best of our knowledge, nonoverlapping cohorts have not been studied in the literature to develop prediction models or analyze the temporal influence of clinical features on kidney transplant outcomes.

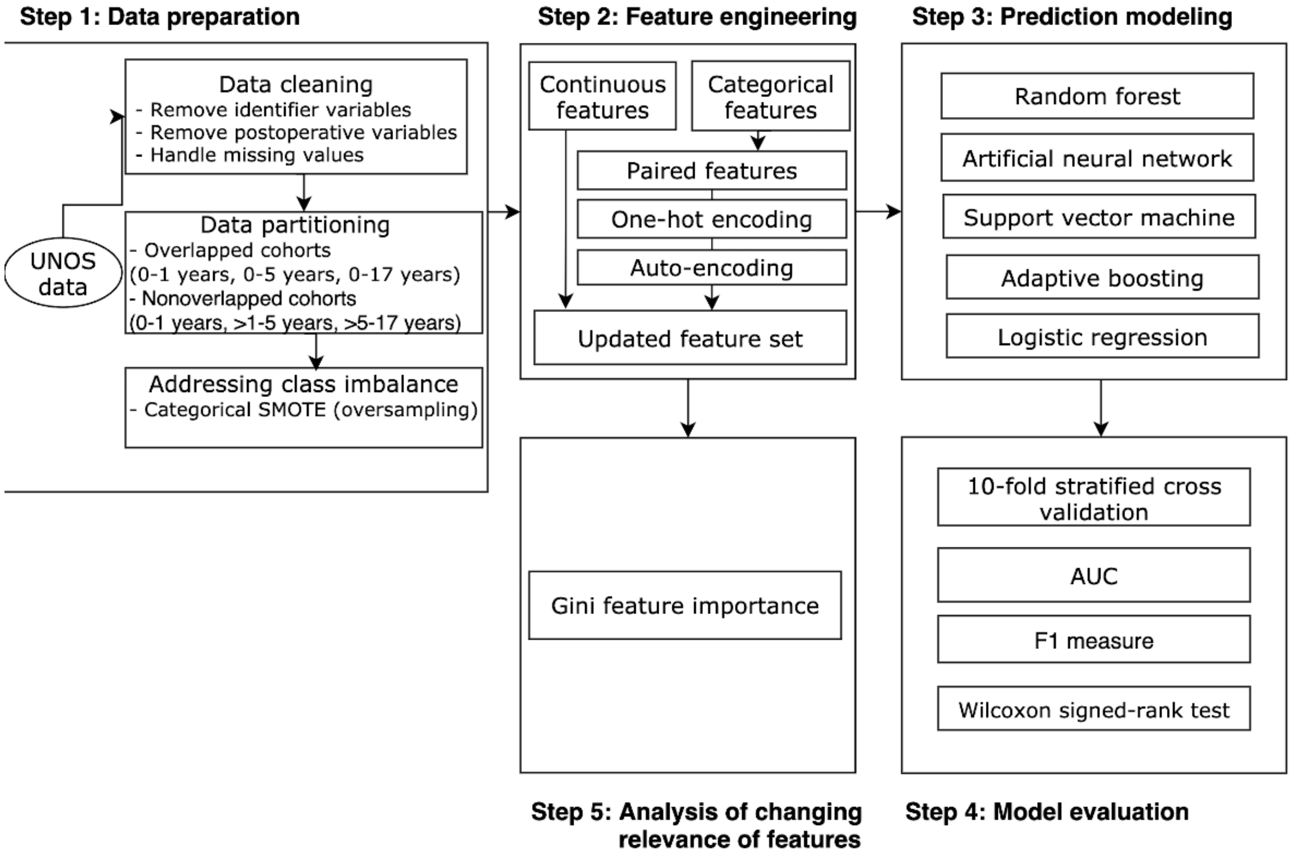
This study is organized into five major sections: *Methods* presents the study's methodology; *Results* presents the results of the prediction; and *Discussion* discusses the significance of clinical features toward graft status prediction across different time cohorts and offers a conclusion and future research directions.

Methods

Overview

To predict graft survival over time and to analyze the influence of clinical features on graft survival, our data analytics methodology (Figure 1) comprised data preparation, feature engineering, prediction modeling, model evaluation, and analysis of changing relevance of features.

Figure 1. Overview of our data analytics methodology. AUC: area under the curve; SMOTE: synthetic minority oversampling technique; UNOS: united network of organ sharing.


Step 1: Data preparation
Data cleaning
- Remove identifier variables
- Remove postoperative variables
- Handle missing values
UNOS data
Data partitioning
- Overlapped cohorts
(0-1 years, 0-5 years, 0-17 years)
- Nonoverlapped cohorts
(0-1 years, >1-5 years, >5-17 years)
Addressing class imbalance
- Categorical SMOTE (oversampling)
Step 2: Feature engineering
Continuous features
Categorical features
Paired features
One-hot encoding
Auto-encoding
Updated feature set
Step 3: Prediction modeling
Random forest
Artificial neural network
Support vector machine
Adaptive boosting
Logistic regression
Gini feature importance
10-fold stratified cross validation
AUC
F1 measure
Wilcoxon signed-rank test
Step 5: Analysis of changing relevance of features
Step 4: Model evaluation

Data Description

This study used data from the Scientific Registry of Transplant Recipients (SRTR). The SRTR data system includes data on all donors, wait-listed candidates, and transplant recipients in the United States, submitted by the members of the Organ Procurement and Transplantation Network. The Health Resources and Services Administration and the US Department of Health and Human Services provided an overview of the activities of the Organ Procurement and Transplantation Network and SRTR contractors.

The data set provided pretransplant clinical features and outcomes of 277,316 kidney transplants between 2000 and 2017. Survival was reported in terms of graft outcome and patient status. For the purposes of this study, graft failure was defined as (1) graft loss or (2) death with a functioning graft.

We analyzed the data and used only complete cases (ie, no missing feature values), which comprised a total of 52,827 kidney transplants. Table 1 provides a list of the included clinical features and their descriptions used in our experiments.

Table 1. List of clinical features used to train the prediction models.

Feature description	Data type	Abbreviation
Peak panel reactive antibody	Continuous	PKPRA
Type of transplant	Categorical	REC_TX_PROCEDURE
Any previous kidney transplant	Categorical	PREVKI
Donor age	Continuous	DAGE
Donor height	Continuous	DHT100
Recipient height	Continuous	RHT2100
Donor weight	Continuous	DWT
Recipient weight	Continuous	RWT2
Donor creatinine level	Continuous	DONCREAT
Expanded criteria donor	Categorical	ECD
Donation after cardiac death	Categorical	DCD
Donor hypertension	Categorical	DHTN2
Recipient hypertension	Categorical	RHTN
Recipient BMI	Continuous	RBMI2
Donor BMI	Continuous	DBMI
Cold ischemia time	Continuous	CIT
Recipient age	Continuous	RAGETX
Number of HLA antigen mismatches (paired)	Categorical	HLAMM
Functional status of the recipient	Categorical	FUNCTSTAT
Donor-recipient sex (paired)	Categorical	DRSEX
Donor-recipient race (paired)	Categorical	DRRACE
Donor-recipient age (paired)	Categorical	DRAGE
Recipient cardiovascular disease	Categorical	RCVD
Donor hepatitis C virus	Categorical	DHCV
Recipient peripheral vascular disease	Categorical	RPVD
Donor race	Categorical	DRACESIMP
Recipient race	Categorical	RRACESIMP
Recipient malignancy	Categorical	RMALIG
Years on dialysis pretransplant	Continuous	VINTAGE
Donor diabetes	Categorical	DDM
Preemptive transplant	Categorical	PREEMPTIVE
Recipient diabetes	Categorical	RDM2
Recipient coronary artery disease	Categorical	RCAD
Simplified ESRD[a] diagnosis	Categorical	ESRDDXSIMP
Donor-recipient CMV[b] (paired)	Categorical	DRCMV
Donor-recipient height difference	Categorical	AHD1
Donor-recipient weight difference	Categorical	DRWT

[a] ESRD: end-stage renal disease.

[b] CMV: cytomegalovirus.

Data Preparation

Data preparation for learning the ML-based prediction models consisted of data cleaning, partitioning the data set into temporal cohorts, and addressing class imbalances.

Data Cleaning

Data cleaning involved removing (1) all patient identifying features (such as transplant ID, donor ID, and patient ID) [23], (2) post- and intraoperative features (such as delayed graft function and warm ischemia time) as we focused on pretransplant features to predict outcomes [23], (3) living donors [15,26], (4) recipients below the age of 18 years [14,27,28], and (5) all sequential and en bloc transplants, both of which are atypical procedures that would not broadly apply to most deceased donor situations. These exclusion criteria were suggested by domain experts and also noted in prior studies [9,28].

Data Partitioning Into Temporal Cohorts

Given the longitudinal data set, we generated two distinct data sets using traditional *overlapping* and our novel *nonoverlapping* patient stratification approaches to partition the data set into three temporal cohorts representing graft status at 1 year, >1-5 years, and 5-17 years (Figures 2 and 3).

Figure 2. Derivation of the overlapped cohorts.

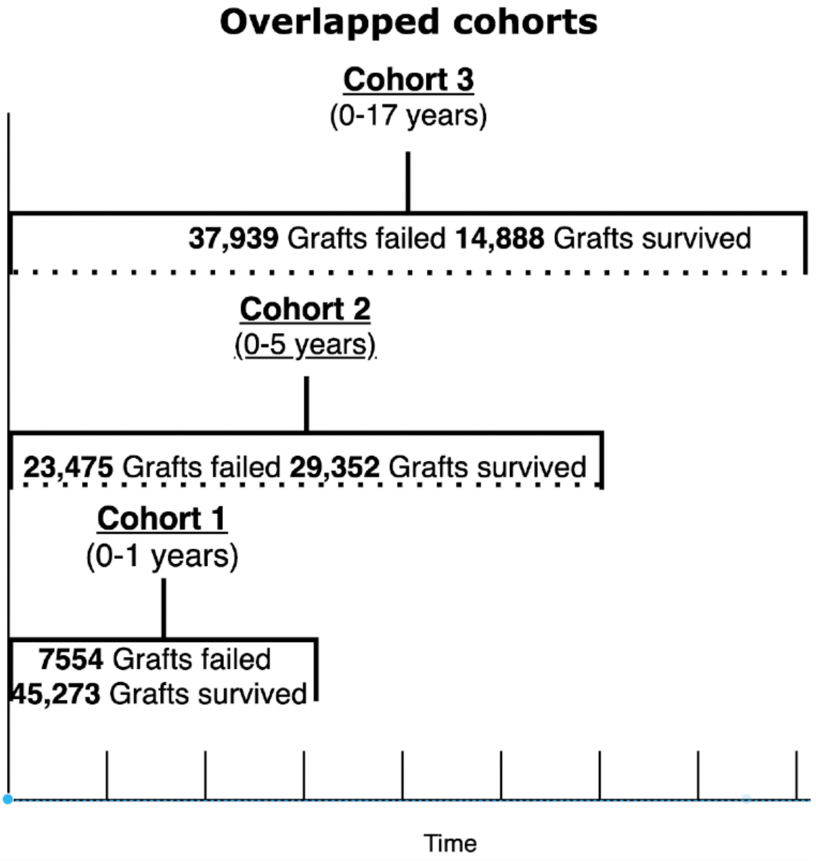
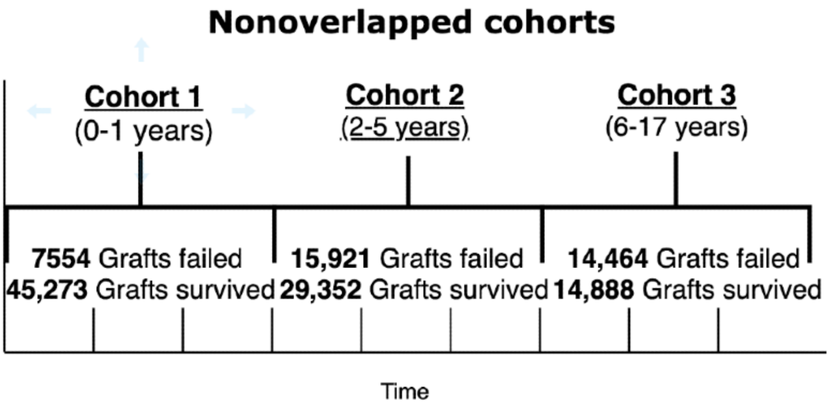

Overlapped cohorts
Cohort 3
(0-17 years)
37,939 Grafts failed 14,888 Grafts survived
Cohort 2
(0-5 years)
23,475 Grafts failed 29,352 Grafts survived
Cohort 1
(0-1 years)
7554 Grafts failed
45,273 Grafts survived
Time

Figure 3. Derivation of the nonoverlapped cohorts.


Nonoverlapped cohorts
Cohort 1
(0-1 years)
Cohort 2
(2-5 years)
Cohort 3
(6-17 years)
7554 Grafts failed
45,273 Grafts survived
15,921 Grafts failed
29,352 Grafts survived
14,464 Grafts failed
14,888 Grafts survived
Time

The overlapping patient stratification approach (used in previous graft status prediction studies) provides a cumulative analysis of graft outcomes up to a specific time point. In our study, the overlapping data stratification resulted in the following three cohorts: cohort 1, spanning from year$_0$ to year$_1$, which reported graft outcomes (ie, graft failure or survival) during this period; cohort 2, which reported graft status from year$_0$ to year$_5$ and overlapped with cohort 1 such that it included the patients in cohort 1; and cohort 3, which reported graft status from year$_0$ to year$_{10+}$, thus overlapping with both earlier cohorts. As per the overlapping approach, a graft failure in the preceding cohort was also counted in the proceeding cohort.

Our nonoverlapping patient stratification approach yielded three cohorts: cohort 1, spanning year$_0$ to year$_{1,}$ reporting graft outcomes in this period; cohort 2, spanning year$_1$ to year$_5$ and reporting graft outcomes only in this period, resulting in the exclusion of graft failures reported in cohort 1 and only reporting the graft outcomes of patients who survived cohort 1; and cohort 3, spanning year$_5$ to year$_{10+}$, reporting the graft outcomes of patients who survived cohort 2. In a nonoverlapping cohort approach, there was no looking back beyond the cohort's starting point, as such a graft failure in the preceding cohort was not counted in the proceeding cohort.

When partitioning the data into cohorts, we accounted for the presence of censored data, that is, the lack of information about the occurrence of an adverse event for a surviving patient. There is no concrete method to determine survivors when confronted with censored data. We initially assumed that those patients who did not fail in a certain cohort could be presumed as survivors. However, this assumption led to two problems: (1) it included censored patients who might have experienced graft failure during the study, and (2) it led to a severe class imbalance between the graft failure and surviving patients. To overcome these problems, we took a two-phase heuristic approach to remove the censored observations to identify *survivors* in each cohort. First, we removed all the censored observations from the cohort being analyzed and labeled all the remaining instances as survived. For instance, the censored data that were removed for cohort 2 were all instances with a missing outcome by the end of cohort 2. The remaining instances were considered to have survived. The first phase of our approach reduced the number of censored observations, but the surviving observations were still relatively high compared with the graft failure cases in each cohort. In the second phase, we applied the approach of Topuz et al [15] to further refine our surviving class by removing all the instances that were deemed as surviving for less than 8 years from the date of transplant. This two-phase approach to account for censored data is summarized in Equation 1 below, where we estimated the number of survivors in each cohort.

The first part of this equation illustrates the first phase of the proposed approach. The *i* in the equation is the ending year of our defined cohorts, that is, 1, 5, and 17. The equation first calculates the total number of graft failures that occurred after the end of the cohort. This fraction of instances was considered as confirmed survivors for the cohort under analysis. The second part of the equation deals with the second phase of the approach. It attempts to identify the potential survivors from the censored data by removing all the observations that did not have a graft failure within 8 years (2920 days) following a transplant. Table 2 shows the patient distribution across the three time cohorts.

Table 2. Number of failed and survived transplants in overlapped and nonoverlapped cohorts.

Cohort	Overlapping			Nonoverlapping		
	Count, n	Failed, n (%)	Survived, n (%)	Count, n	Failed, n (%)	Survived, n (%)
1	52,827	7554 (14.3)	45,273 (85.7)	52,827	7554 (14.3)	45,273 (85.7)
2	52,827	23,475 (44.44)	29,352 (55.56)	45,273	15,921 (35.17)	29,352 (64.83)
3	52,827	37,939 (71.82)	14,888 (28.18)	29,352	14,464 (49.28)	14,888 (50.72)

Addressing Class Imbalance

Our data set had two outcomes: the presence or absence of graft failure. There was a significant class imbalance whereby the *graft failure* had a significantly lower number of instances compared with *graft survival*. Techniques such as Synthetic Minority Oversampling Technique (SMOTE) and random undersampling have been widely used in the literature [18,29] for class imbalance. We applied SMOTE for Nominal and Continuous features [30], which is a variant of SMOTE specifically developed to handle a mix of categorical and numerical data, on all cohorts to achieve a reasonable class balance. To balance the minority class (ie, mostly *graft failure*), we would need to generate 600% additional synthetic samples (at least for cohort 1), which would have led to overfitting. Therefore, we doubled our minority class to achieve a workable class balance to prevent the overfitting of the classification models. As cohort 2 of overlapped stratification and cohort 3 of nonoverlapped stratification had a class balance, they were not considered for oversampling.

Feature Engineering

This step involved both the removal and construction of features with the intent to reduce the dimensionality of the feature space.

Paired Features

A set of paired features was constructed by combining the related features. Typically, graft predictions use discrete individual donor and recipient features. We examined the underlying correlation between the donor and recipient features and paired the highly related features to generate new *paired features*. The following four donor and recipient features were generated: sex, age, CMV (cytomegalovirus), and race. In addition, three types of HLA Antigen Mismatch features—ie, HLA Antigen Mismatch at the A Locus, HLA Antigen Mismatch at the B Locus, and HLA Antigen Mismatch at the

DR Locus were also combined into a single HLA Antigen Mismatch feature.

One-Hot Encoding

We transformed the categorical features into multiple dummy features to make them compatible with the stacked autoencoders, which cannot process categorical features. In addition, it was also a necessary operation because of the functional constraint of the scikit-learn library [31].

Stacked Autoencoders

Finally, we used 86 transformed categorical features as inputs to the stacked autoencoders for feature reduction to subsequently train the ML prediction models. Continuous features were also initially considered as a part of the input vector to stacked autoencoders (Table S1 in Multimedia Appendix 1), but preliminary model training returned better results with pristine continuous features; hence, no modification was performed for them while training the prediction models. It should be noted that the resultant features from the stacked autoencoders were only used for training the prediction models and not for the analysis of the changing relevance of features over time.

After testing with different configuration settings provided in the Keras framework [32], the stacked autoencoders were set up as a 13-layer architecture consisting of 12 dense layers and one dropout layer set at the very beginning of the network with a dropout rate of 0.05. The sigmoid activation function was used throughout the dense layers, with Adam as the optimizer and binary cross entropy as the loss function with 500 epochs and 700-900 batch sizes. The middle layer of the autoencoder was finally trained with 12 neurons and 30 neurons for cohort 1 and the remaining cohorts, respectively. The feature space was reduced by more than 50 dummy features.

Learning Classification-Based Graft Survival Prediction Models

Overview

Prediction was pursued as a binary classification problem, where the prediction output represents the graft outcome for a given patient in terms of the class label, graft failure or survived. We investigated four different ML-based classification models for each time cohort (ie, cohorts 1-3). Given that logistic regression (LR) has been widely used in prior studies to develop graft prediction models [29,33], we trained an LR model as a baseline to compare the predictive performance of our ML prediction models.

All classification models were trained using a 10-fold stratified cross-validation training approach. The stratification ensured that outcome class ratio in each fold is maintained to avoid any sampling bias that may affect the classification results. We mainly used the scikit-learn library [31] to train the below-mentioned classification models with the parameter settings listed in Table 3. Hyperparameters were optimized using a random search. The different parameters that were tested during the random search are provided in Table S2 in Multimedia Appendix 1.

Table 3. Algorithmic settings for the classifiers.

Method, Hyperparameter	Values
RF[a]	
Number of estimators	200
Class weight	Balanced
Criterion	Gini
Maximum depth	9
Minimum samples split	2 for cohort 1; 3 for the rest
Maximum features	14
Support vector machine	
C[b]	50
Gamma	Auto, scale
Decision function shape	One versus rest
Kernel	Radial
Artificial neural network	
Solver	Adam
Learning rate	Adaptive
Activation	Logistic
Alpha	1e-2,1e-6
Hidden layers	*4*: 70, 35, 30, 15; *5*: 60, 30, 30, 15, 10
Adaptive boosting	
Base learner	RF
Number of estimators	401
Learning rate	1
Algorithm	Samme.R
Logistic regression	
Penalty	l2
C	10
Class weight	Balanced
Max iteration	1000
Solver	Sag

[a] RF: random forest.

[b] C: regularization parameter.

Random Forest

Random forest (RF) was used as both a standalone classifier and a base learner for the adaptive boosting (AdaBoost) algorithm. It has been widely used to predict survival data [26,27].

AdaBoost

The AdaBoost algorithm was applied to two weak learners, RF and LR. The study by Thongkam et al [34] used this algorithm on a breast cancer data set, where it outperformed all single classifiers. In our experiments, LR did not perform well; therefore, we did not pursue it. RF, with the optimized hyperparameters, was used to train the boosting classifier with a number of estimators and learning rates.

Artificial Neural Network

A backpropagation algorithm was used to train a neural network–based binary classifier. Generally, artificial neural networks (ANNs) perform well on survival data sets [30,35].

Support Vector Machines

Classification models using support vector machines (SVMs) have been applied to predict survival data [29,31,32]. To train the SVM, we experimented with different kernels, that is, linear, radial, sigmoid, and polynomial kernels. The linear and sigmoid kernels provided the lowest prediction scores; therefore, we did not use them further. A polynomial kernel with degree 2 yielded suboptimal results, and the SVM model could not converge for degree 3. The radial basis kernel was the most effective for learning the classification model.

Calculating Feature Importance Over Time

The nonoverlapped time cohorts were used to calculate the feature importance scores to understand the changing relevance of features over time. We calculated these scores by training an RF classifier on the complete data set. The scores were calculated using Gini. Feature influence scores were used to understand the effect of features over the three cohorts.

Results

Overview

Below, we present the prediction performance of the four ML classifiers using both overlapped and nonoverlapped cohorts. As LR has been extensively used to predict time-to-event in organ transplant studies [16,29], it is used as a comparator classifier to the ML-based classifiers. The prediction performance of each of the best-trained classifiers (SVM, RF, AdaBoost, ANN, and LR), covering the three different time-to-event periods for both the original and reduced feature sets, were evaluated using 10-fold stratified cross validation for both overlapped and nonoverlapped cohorts. The results of each classifier were examined using the area under the curve (AUC) and F1 scores. The AUC score was used as the main performance evaluation metric to select the best model in each cohort and to make comparisons with similar studies. For our purpose, the ideal prediction model provides the best accuracy for graft failure. Therefore, to further substantiate the selection of the best model, we also evaluated the F1 score for graft failures. In cases where the AUC score was the same for different models, preference was given to the model with the highest F1 score.

Analysis of Feature Engineering

Table 4 presents the results of feature engineering, whereby the prediction scores of all classifiers were obtained using both the original feature set and the reduced set. The reduced set consists of original continuous features and latent features returned by the stacked autoencoders. Cohort 1 for overlapped and nonoverlapped cohorts was the same; hence, the results were presented only once to avoid unnecessary duplication.

Table 4. Area under the curve comparison—all features with auto-encoded features.

Cohort	Overlapped		Nonoverlapped	
	All features (%)	Continuous+auto-encoded features (%)	All features (%)	Continuous+auto-encoded features (%)
Cohort 1				
SVM[a]	80	82	N/A[b]	N/A
AdaBoost[c]	76	78	N/A	N/A
RF[d]	68	70	N/A	N/A
ANN[e]	62	61	N/A	N/A
LR[f]	62	62	N/A	N/A
Cohort 2				
SVM	63	66	53	53
AdaBoost	67	69	64	60
RF	62	65	65	67
ANN	62	62	62	62
LR	62	62	64	61
Cohort 3				
SVM	73	80	68	65
AdaBoost	76	81	68	64
RF	72	75	68	66
ANN	73	72	68	65
LR	69	69	62	64

[a] SVM: support vector machine.

[b] N/A: not applicable.

[c] AdaBoost: adaptive boosting.

[d] RF: random forest.

[e] ANN: artificial neural network.

[f] LR: logistic regression.

The AUC scores (Table 4) show that prediction models for overlapped cohorts trained with auto-encoded features improved the prediction performance as compared with the prediction models trained using the original feature set. However, it was the opposite for nonoverlapped cohorts. Interestingly, the traditional approach of overlapped cohorts performed better with both the original and reduced feature sets compared with the nonoverlapped cohorts. Except for RF in cohort 2 of nonoverlapped cohorts, which showed slightly better performance (67%) when compared with its overlapped counterpart (65%), all other prediction models had better AUC scores with the traditional overlapping cohort approach. Therefore, for further analysis, we proceeded with overlapping cohorts only.

Although ANN and LR (the baseline model) showed no significant improvement across all three cohorts, the results confirmed the effectiveness of our deep learning architecture of stacked autoencoders for feature selection. For the subsequent prediction modeling analysis, we used the reduced feature set.

Analysis of Prediction Performance of ML Models

Table 5 presents the prediction performance of the classifiers for each cohort in terms of AUC, F1 scores, recall, and precision, with SD for the 10-fold classification.

Table 5. Prediction performance of the machine learning classifiers across three different temporal cohorts using the overlapped patient stratification[a].

Cohort	Auto-encoded feature set			
	AUC[b] (%), mean (SD)	F1 (%), mean (SD)	Recall (%), mean (SD)	Precision (%), mean (SD)
Cohort 1				
SVM[c]	*82 (0.01)*	*61 (0.01)*	*49 (0.01)*	*90 (0.01)*
AdaBoost[d]	78 (0.01)	56 (0.01)	95 (0.01)	35 (0.01)
RF[e]	70 (0.009)	45 (0.001)	47 (0.01)	41 (0.01)
ANN[f]	61 (0.01)	5 (0.001)	42 (0.01)	6 (0.004)
LR[g]	62 (0.008)	39 (0.009)	58 (0.01)	29 (0.04)
Cohort 2				
SVM	66 (0.006)	53 (0.01)	55 (0.01)	60 (0.01)
AdaBoost	*69 (0.01)*	*63 (0.01)*	*64 (0.003)*	*63 (0.004)*
RF	65 (0.009)	62 (0.01)	62 (0.01)	61 (0.01)
ANN	63 (0.007)	60 (0.04)	55 (0.09)	60 (0.01)
LR	62 (0.008)	59 (0.009)	58 (0.01)	60 (0.004)
Cohort 3				
SVM	80 (0.005)	83 (0.003)	76 (0.003)	96 (0.003)
AdaBoost	*81 (0.01)*	*81 (0.004)*	*76 (0.003)*	*86 (0.01)*
RF	75 (0.008)	75 (0.006)	75 (0.01)	73 (0.01)
ANN	72 (0.007)	68 (0.005)	81 (0.03)	69 (0.01)
LR	69 (0.001)	77 (0.009)	70 (0.01)	70 (0.001)

[a] Italics show the classifiers with the highest performance among the three cohorts.

[b] AUC: area under the curve.

[c] SVM: support vector machine.

[d] AdaBoost: adaptive boosting.

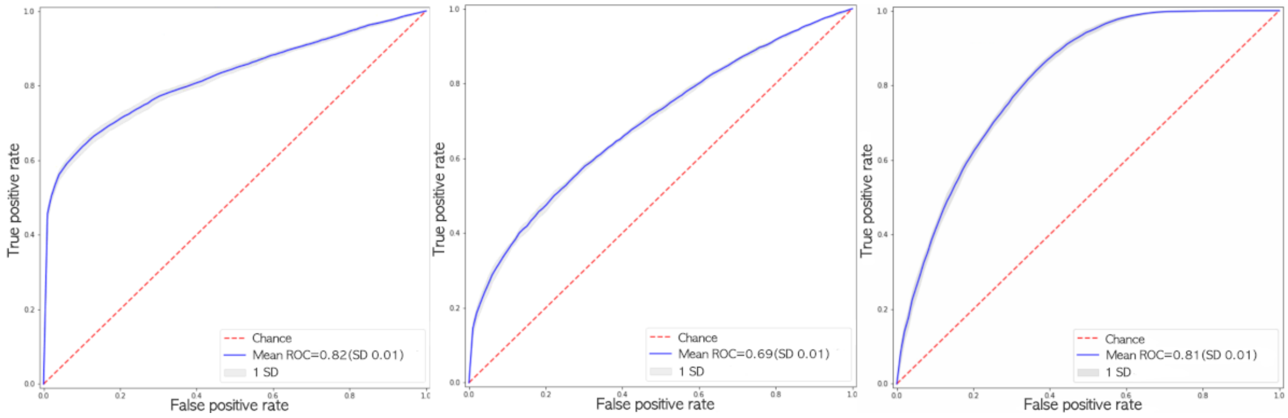
[e] RF: random forest.

[f] ANN: artificial neural network.

[g] LR: logistic regression.

The classifiers performed differently across the three cohorts—SVM offered the highest prediction performance for short-term predictions, that is, for cohort 1, whereas AdaBoost offered the highest performance for the remaining cohorts. The SD across the different folds was nominal, confirming the stability of the classifiers. Figure 4 shows the receiver operating characteristic curves for the best models from each cohort.

Figure 4. Receiver operating characteristic curves for support vector machine, adaptive boosting, and adaptive boosting for the three cohorts, respectively (left to right). AUC: area under the curve; ROC: receiver operating characteristic.



To further investigate the prediction efficacy of the ML-based classifiers, we evaluated the prediction performance of the best-performing classifier for all three cohorts by testing the prediction of graft failure events by a classifier trained for a specific cohort with data from other cohorts, that is, testing the classifier for cohort 2 with randomly selected data from cohorts 1 and 3. The underlying assumption is that the classifier should not produce good prediction results for data from other cohorts. As this evaluation considers survivors across progressive cohorts, we used the F1 score to measure prediction performance. A sound prediction model for cohort 2 will give a high graft failure prediction score for data from cohort 1 but a low prediction score for data from cohort 3, the rationale being that the overlapping cohort 2 classifier is trained on graft failure cases in both cohorts 1 and 2. Therefore, the prediction model for cohort 2 should give a high prediction score for predicting graft failures from $year_0$ to $year_5$, but when applied to cohort 3, the cohort 2 prediction model would be unable to predict graft failure as it has not been trained on cohort 3 data. Table 6 provides the cross-cohort prediction scores for the best classifiers for each cohort.

Table 6. Prediction performance (F1 scores) for cross-cohort predictions using overlapped cohorts.

Model	Cohort 1	Cohort 2	Cohort 3
SVM[a] (cohort 1)	0.6	0.42	0.29
AdaBoost[b] (cohort 2)	0.79	0.87	0.58
AdaBoost (cohort 3)	0.72	0.75	0.87

[a] SVM: support vector machine.

[b] AdaBoost: adaptive boosting.

Results of Wilcoxon Signed-Rank Test

To determine if the prediction differences between the different models were statistically significant, we used the Wilcoxon signed-rank test to compare the scores between different models. Because the best scores in each cohort were usually produced by SVM and AdaBoost models, the Wilcoxon signed-rank test was conducted with each combination of these models with the other models.

Table 7 shows the results based on the F1 score, and the *P* values between the models were quite small and less than the threshold value of *P*=.05, confirming that the performance difference is statistically significant.

Table 7. The results for Wilcoxon signed-rank test (F1).

Cohort	*P* value (F1)				
	SVM[a]-AdaBoost[b,c]	SVM-ANN[d,e]	SVM-RF[f,g]	AdaBoost-RF[h]	AdaBoost-ANN[i]
Cohort 1	.003	.003	.003	.003	.003
Cohort 2	.003	.003	.003	.003	.03
Cohort 3	<.001	<.001	<.001	<.001	<.001

[a] SVM: support vector machine.

[b] AdaBoost: adaptive boosting.

[c] H_0 (null hypothesis): SVM=AdaBoost; H_a (alternative hypothesis): SVM≠AdaBoost.

[d] ANN: artificial neural network.

[e] H_0: SVM=EANN; H_a: SVM≠EANN.

[f] RF: random forest.

[g] H_0: SVM=RF; H_a: SVM≠RF.

[h] H_0: AdaBoost=RF; H_a: AdaBoost≠RF.

[i] H_0: AdaBoost=ANN; H_a: AdaBoost≠ANN.

Analysis of the Influence of Clinical Features Over Time Toward Graft Status Prediction

Overview

The second objective of this research is to analyze the influence of clinical features on the prediction of graft survival over different periods. The intent was to understand the factors responsible for graft survival at different periods after transplant. The nonoverlapped cohorts (0-1 years, >1-5 years, and >5-17 years following a transplant) were used to ensure that there was no cascading influence of the features over time. For comparison purposes, we also examined the feature importance for overlapping cohorts. The feature importance scores represent the relative importance of the feature among all features, that is, the total of all the features' importance scores add up to 100%; hence, if one feature gains a higher importance score, it will be at the expense of the importance score of other features.

Figures 5 and 6 illustrate the individual feature importance scores across all the nonoverlapped and overlapped cohorts, respectively.

Figure 5. Changing relevance of features based on nonoverlapped time cohorts.

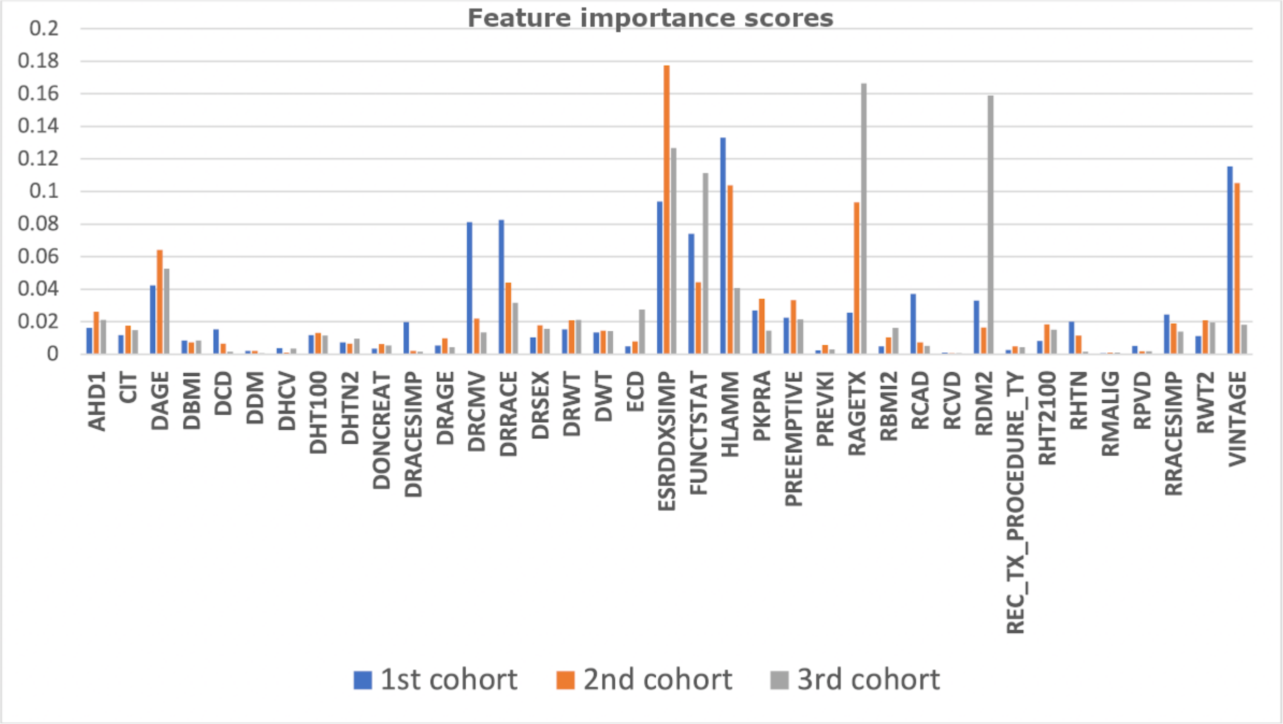
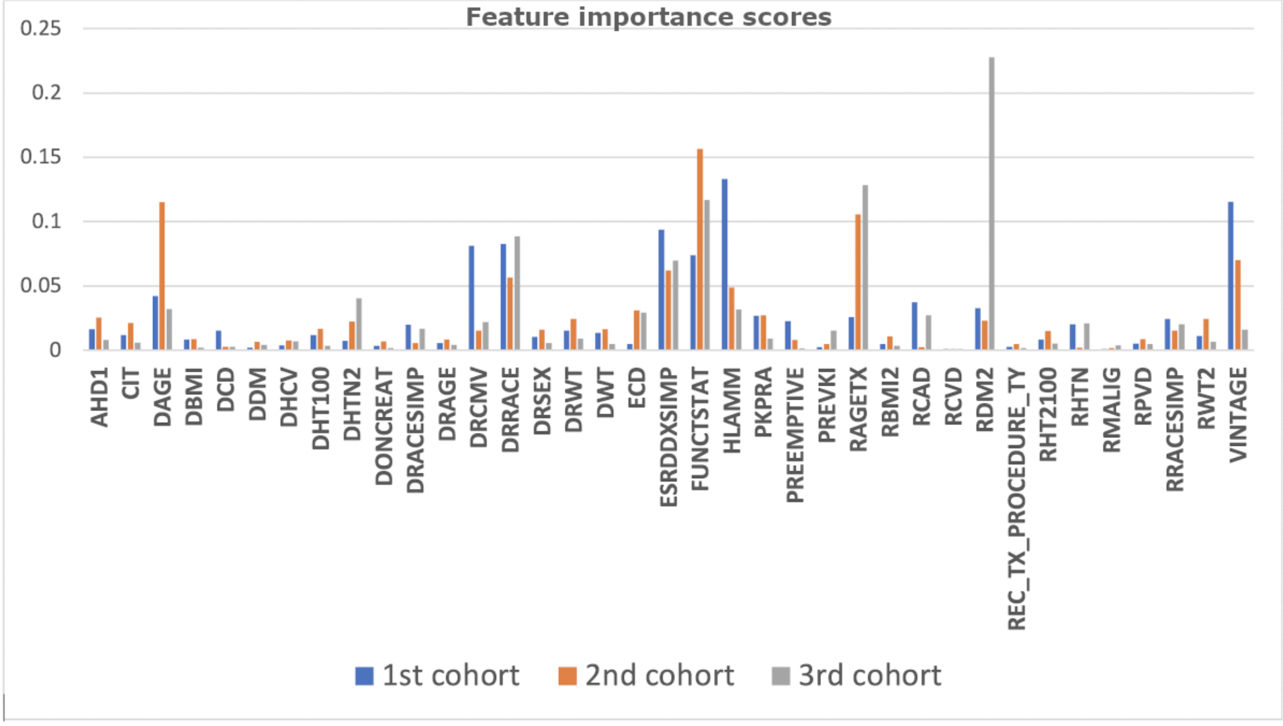


Figure 6. Changing relevance of features based on overlapped time cohorts.



In general, the top 10% of the important features remained consistent in both the nonoverlapped and overlapped cohorts; however, we note that the nonoverlapped cohorts identified a larger group of important features. For instance, peak panel reactive antibody (Pkpra) and pre-emptive recipient status (Preemptive) had negligible importance in overlapping cohorts but were important during the 2-5 years and 6-17 years in the nonoverlapping cohorts. Table 8 shows the feature importance across the three cohorts.

Table 8. Ranking of the top-10 features across the time cohorts with feature importance scores[a].

Rank	Cohort 1, feature, relative score (%)	Cohort 2		Cohort 3	
		Feature, relative score (%)	Importance (%), rank change	Feature, relative score (%)	Importance (%), rank change
1	HLAMM[b] (13)	ESRDDXSIMP[c] (18)	*+100, +2*	RAGETX[d] (16)	+77, *+3*
2	VINTAGE[e] (12)	VINTAGE (11)	–8, *0*	RDM2[f] (16)	>+100, >*+10*
3	ESRDDXSIMP (9)	HLAMM (10)	–23, *–2*	ESRDDXSIMP (13)	–27, *–2*
4	DRCMV[g] (8)	RAGETX (9)	+78, *+6*	FUNCTSTAT[h] (12)	>+100, *+3*
5	DRRACE[i] (8)	DAGE[j] (6)	50, *+2*	DAGE (5)	–16, *0*
6	FUNCTSTAT (7)	DRRACE (4)	–100, *–1*	DRRACE (3)	–25, *0*
7	DAGE (4)	FUNCTSTAT (4)	–75, *–1*	ECD[k] (3)	>+100, >*+3*
8	RCAD[l] (4)	PKPRA[m] (3)	+50, >*+3*	VINTAGE (2)	>–100, *–6*
9	RDM2 (3)	PREEMPTIVE[n] (3)	+50, >*+1*	PREEMPTIVE (2)	–33, *0*
10	RAGETX (2)	DRCMV (2)	–75, *–6*	RWT2 (2)	0, *0*
Rest	Rest (31)	Rest (30)	Rest (30)	Rest (26)	Rest (26)

[a] Importance (%) and rank change is shown in italics.

[b] HLAMM: HLA antigen mismatch.

[c] ESRDDXSIMP: simplified end-stage renal disease diagnosis.

[d] RAGETX: recipient age.

[e] VINTAGE: number of years on dialysis before transplant.

[f] RDM2: recipient diabetes status.

[g] DRCMV: donor-recipient cytomegalovirus.

[h] FUNCTSTAT: functional status of the recipient.

[i] DRRACE: donor-recipient race.

[j] DAGE: donor age.

[k] ECD: expanded criteria donor.

[l] RCAD: recipient coronary artery disease.

[m] PKPRA: peak panel reactive antibody.

[n] PREEMPTIVE: pre-emptive transplant.

[o] RWT2: recipient weight.

Below, we analyze the importance of features in each cohort and show the influence of features over time using nonoverlapping cohorts.

Feature Importance for Cohort 1

According to the top features shown in Table 8, HLAMMs and the number of years on dialysis before transplant (VINTAGE) were the most important features with a relative importance of over 10%. This observation has been confirmed in other studies [32,36]. Donor-recipient CMV status, donor-recipient race, end-stage renal disease diagnosis (ESRDDXSIMP), and functional status of the recipient were ranked as having medium importance with a relative score between 5% and 10%. Donor-recipient race pairs and donor-recipient CMV pairs were noted to have more predictive influence in cohort 1 than in the other two cohorts.

Feature Importance for Cohort 2

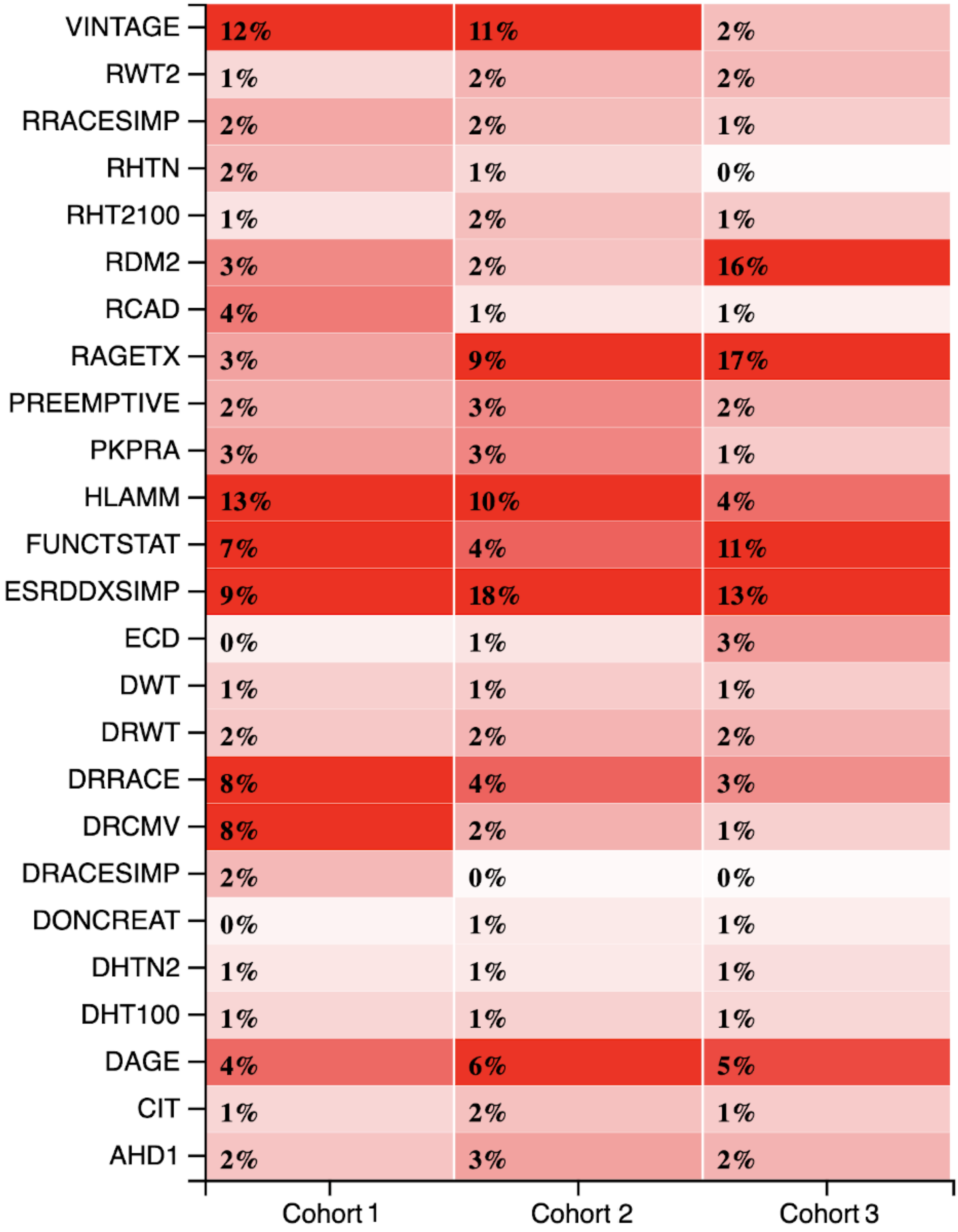
Both HLAMMs and VINTAGE remained highly important in cohort 2. In addition, ESRDDXSIMP was noted as a highly important feature. Interestingly, we note that few features, such as donor age and recipient age, were rather insignificant in cohort 1 but were noted to be significant in both cohort 2 and further in cohort 3.

Feature Importance for Cohort 3

ESRDDXSIMP showed a relative downward trend; however, it remained a highly significant feature. Unlike earlier cohorts, HLAMMs and VINTAGE were noted to not maintain their importance in the long term, whereas the recipient's status of diabetes was noted to be the most important feature, along with recipient age and their functional status. Donor age was noted to maintain a medium importance score between 5% and 10%.

Figure 7 presents a heat map of the importance score to illustrate the changing influence of the top 25 features across the three cohorts. In addition to the top 10 features (Table 8), the heat map details the contribution of relatively less important features. It was interesting to see that few features (such as donor weight, recipient weight, and donor hypertension) had static importance across the three cohorts. This indicates that although these features were not deterministic for the cohorts, they possessed a certain value for the prediction models.

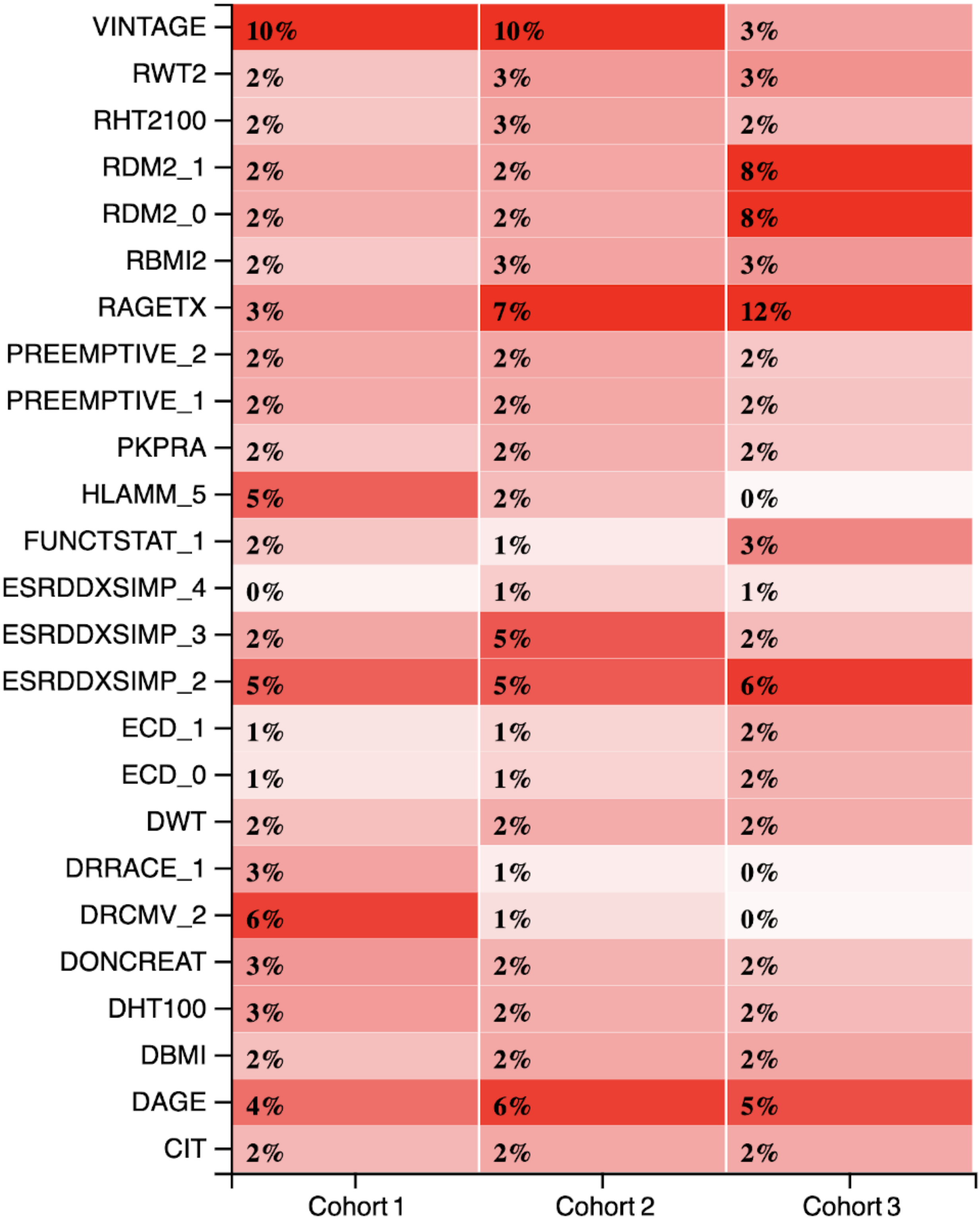
Figure 7. Changing relevance of top 25 features over the three cohorts. AHD1: donor-recipient height difference; CIT: cold ischemia time; DAGE: donor age; DHT100: donor height; DHTN2: donor hypertension; DONCREAT: donor creatinine level; DRACESIMP: donor race; DRCMV: donor-recipient cytomegalovirus; DRRACE: donor-recipient race; DRWT: donor-recipient weight difference; DWT: donor weight; ECD: expanded criteria donor; ESRDDXSIMP: simplified end-stage renal disease diagnosis; FUNCTSTAT: functional status of the recipient; HLAMM: number of HLA mismatches; PKPRA: peak panel reactive antibody; PREEMPTIVE: preemptive transplant; RAGETX: recipient age; RCAD: recipient coronary artery disease; RDM2: recipient diabetes; RHT2100: recipient height; RHTN: recipient hypertension; RRACESIMP: recipient race; RWT2: recipient weight; VINTAGE: number of years on dialysis before transplantation.



	Cohort 1	Cohort 2	Cohort 3
VINTAGE	12%	11%	2%
RWT2	1%	2%	2%
RRACESIMP	2%	2%	1%
RHTN	2%	1%	0%
RHT2100	1%	2%	1%
RDM2	3%	2%	16%
RCAD	4%	1%	1%
RAGETX	3%	9%	17%
PREEMPTIVE	2%	3%	2%
PKPRA	3%	3%	1%
HLAMM	13%	10%	4%
FUNCTSTAT	7%	4%	11%
ESRDDXSIMP	9%	18%	13%
ECD	0%	1%	3%
DWT	1%	1%	1%
DRWT	2%	2%	2%
DRRACE	8%	4%	3%
DRCMV	8%	2%	1%
DRACESIMP	2%	0%	0%
DONCREAT	0%	1%	1%
DHTN2	1%	1%	1%
DHT100	1%	1%	1%
DAGE	4%	6%	5%
CIT	1%	2%	1%
AHD1	2%	3%	2%

Analysis of the values of categorical features provided novel insights into the influence of a feature. Figure 8 presents a heat map of the importance of the value of the categorical features generated after transforming them into dummy features.

Figure 8. Changing relevance of top 25 features (including dummy features) over the three cohorts. CIT: cold ischemia time; DAGE: donor age; DBMI: donor BMI; DHT100: donor height; DONCREAT: donor creatinine level; DRCMV_2: Donor positive recipient positive; DRRACE_1: Donor white recipient white; DWT: donor weight; ECD_0: Expanded criteria donor: no; ECD_1: Expanded criteria donor: yes; ESRDDXSIMP_2: End stage renal disease: diabetes mellitus; ESRDDXSIMP_3: End stage renal disease: polycystic kidney disease; ESRDDXSIMP_4: End stage renal disease: hypertension; FUNCTSTAT_1: Functional status of recipient: 100% no complaints; HLAMM _5: Number of human leukocyte antigen mismatches: 5; PKPRA: peak panel reactive antibody; PREEMPTIVE_1: Preemptive transplant: yes; PREEMPTIVE_2: Preemptive transplant: no; RAGETX: recipient age; RBMI2: recipient BMI; RDM2_0: Recipient diabetes: no; RDM2_1: Recipient diabetes: yes; RHT2100: recipient height; RWT2: recipient weight; VINTAGE: number of years on dialysis before transplantation.



	Cohort 1	Cohort 2	Cohort 3
VINTAGE	10%	10%	3%
RWT2	2%	3%	3%
RHT2100	2%	3%	2%
RDM2_1	2%	2%	8%
RDM2_0	2%	2%	8%
RBMI2	2%	3%	3%
RAGETX	3%	7%	12%
PREEMPTIVE_2	2%	2%	2%
PREEMPTIVE_1	2%	2%	2%
PKPRA	2%	2%	2%
HLAMM_5	5%	2%	0%
FUNCTSTAT_1	2%	1%	3%
ESRDDXSIMP_4	0%	1%	1%
ESRDDXSIMP_3	2%	5%	2%
ESRDDXSIMP_2	5%	5%	6%
ECD_1	1%	1%	2%
ECD_0	1%	1%	2%
DWT	2%	2%	2%
DRRACE_1	3%	1%	0%
DRCMV_2	6%	1%	0%
DONCREAT	3%	2%	2%
DHT100	3%	2%	2%
DBMI	2%	2%	2%
DAGE	4%	6%	5%
CIT	2%	2%	2%

Discussion

Principal Findings

The cross-cohort prediction results (Table 6) confirm the efficacy of the classifiers—the prediction model for cohort 3 (ie, AdaBoost) correctly offers a high prediction score for data from cohort 1 (72%) and cohort 2 (75%). The prediction model for cohort 2 offers a high prediction score for cohort 1 data (79%) but a low prediction score for cohort 3 (58%) data. The classifier for cohort 1 (ie, SVM) gave low prediction scores for data from cohort 2 (42%) and cohort 3 (29%). Interestingly, the highest prediction score by a cohort-specific classifier was always achieved for data from its respective cohort. The prediction modeling results confirmed that the prediction models were highly sensitive to their respective cohorts.

Comparing Prediction Performance With Prior Studies

We compared the prediction performance of our ML-based prediction models with comparable organ transplant studies that involved similar-sized observations and temporal windows. Table 9 summarizes the findings of the two studies for each cohort. There have been several other studies [19,31,32,35,37,38] to predict the short-term graft status of different organ transplants, but because of their small data set, these do not serve as a meaningful comparison.

Table 9. Prediction scores of similar studies.

Time	Study	Model	Size	Data set	Metric	Score (%)	Our score (%)
1 year	Lin et al [16]	ANN[a] and LR[b]	46,414	UNOS[c]	AUC[d]	73	82
1 year	Dag et al [23]	LR	15,580	UNOS	AUC	63	82
5 years	Tiong et al [39]	Nomogram	20,085	UNOS	C-index[e]	71	69
5 years	Lin et al [16]	ANN	17,856	UNOS	AUC	77	69
7 years	Lin et al [16]	ANN	10,250	UNOS	AUC	82	81
14 years	Luck et al [40]	ANN	46,098	SRTR[f]	C-index	65	81

[a] ANN: artificial neural network.

[b] LR: logistic regression.

[c] UNOS: United Network of Organ Sharing.

[d] AUC: area under the curve.

[e] C-index: concordance index.

[f] SRTR: Scientific Registry of Transplant Recipients.

When comparing our results with prior studies, it is noted that although our cohort 2 prediction performance (ie, graft status prediction over a 5-year period) is lower than that of Lin et al [16], it was based on a much smaller data set that included 10,641 survivals and 7215 failures, whereas we analyzed 23,475 failures and 29,352 survivals. Similarly, Tiong et al [39] analyzed a smaller sample of 20,085 living donor transplant recipients to achieve a concordance index of 71%. Our cohort 3 prediction performance is marginally lower compared with Lin et al [16], who predicted a 7-year graft survival with an 82% AUC score, whereas our cohort 3 prediction model covers a much longer (17 years) temporal window and achieves a comparable prediction score. Using a similar number of transplants, Luck et al [40] achieved a much lower concordance index between 63% and 66% for 14-years graft survival.

Limitations and Future Work

A limitation of our research lies in the removal of censored instances. We removed all successful cases that were censored before 8 years following transplant. Although this type of approach has previously been used, including censored cases is a potential consideration for future analyses.

Conclusions

Understanding the impact of donor and recipient factors that predict short- and long-term kidney transplant allograft survival is important for patients and providers. Kidney transplantation is the optimal form of kidney replacement therapy, but kidney allografts are a limited resource. In addition, the alternative to kidney transplantation (ie, dialysis) is considerably costlier.

In this study, we present an ML-based framework to predict the status of kidney allografts, based on donor-recipient features, over a period of 17 years. We applied ML-based data analysis methods for feature engineering to reduce data dimensionality, develop prediction models for three distinct temporal cohorts, and investigate the changing relevance of clinical features across different temporal cohorts. We introduced the concept of nonoverlapped cohorts to analyze the changing relevance of features in three defined periods. In conclusion, our results emphasize that ML can be effective in predicting graft survival using donor and recipient factors that are routinely collected as part of patient care. As a next step, we plan to incorporate the prediction models into clinical care at the time of allocation; models that best predict short- and long-term kidney graft survival may be used as a pragmatic prognostic tool to aid clinicians in maximizing the best possible matching of donors and recipients while preserving existing allocation rules that are used to promote equity [41].

Acknowledgments

The data reported here have been supplied by the Hennepin Healthcare Research Institute as the contractor for the SRTR. The interpretation and reporting of these data are the responsibility of the authors and in no way should be seen as an official policy of or interpretation by the SRTR or the US Government.

Authors' Contributions

SAAN and SSRA were responsible for the overall data analysis methodology, data analysis using ML algorithms, evaluation of the data analysis results, and writing of the manuscript. KT and AV provided clinical expertise in defining the problem, interpretation of the data, preprocessing of the data and interpretation of the results, providing the data from the data source, and editing the manuscript for clinical clarity and purpose. PCR facilitated the setting up and conducting of data analysis experiments. All authors critically reviewed the manuscript for scientific content and approved the final manuscript for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables.
[DOC File , 24 KB - jmir_v23i8e26843_app1.doc]

References

1. Sullivan CE. Kidney Transplantation. Transplantation 1968;6(5):753. [doi: 10.1097/00007890-196808000-00031]
2. Hall YN, Chertow GM. End stage renal disease. BMJ Clin Evid 2007 Oct 17;2007:1048-1059 [FREE Full text] [Medline: 19450356]
3. Go AS, Chertow GM, Fan D, McCulloch CE, Hsu C. Chronic Kidney Disease and the Risks of Death, Cardiovascular Events, and Hospitalization. N Engl J Med 2004 Sep 23;351(13):1296-1305. [doi: 10.1056/nejmoa041031]
4. Rashidi Khazaee P, Bagherzadeh J, Niazkhani Z, Pirnejad H. A dynamic model for predicting graft function in kidney recipients' upcoming follow up visits: A clinical application of artificial neural network. Int J Med Inform 2018 Nov;119:125-133 [FREE Full text] [doi: 10.1016/j.ijmedinf.2018.09.012] [Medline: 30342680]
5. Schold J, Segev DL. Increasing the pool of deceased donor organs for kidney transplantation. Nat Rev Nephrol 2012 Mar 27;8(6):325-331 [FREE Full text] [doi: 10.1038/nrneph.2012.60] [Medline: 22450438]
6. Conservative Care - The Kidney Foundation of Canada | La Fondation canadienne du rein. URL: https://www.kidney.ca/conservative-care [accessed 2019-11-15]
7. About Chronic Kidney Disease | National Kidney Foundation. URL: https://www.kidney.org/atoz/content/about-chronic-kidney-disease [accessed 2019-11-15]
8. Perl J. Kidney transplant failure: failing kidneys, failing care? Clin J Am Soc Nephrol 2014 Jul;9(7):1153-1155 [FREE Full text] [doi: 10.2215/CJN.04670514] [Medline: 24903388]
9. Vinson A, Kiberd B, Davis R, Tennankore K. Nonimmunologic Donor-Recipient Pairing, HLA Matching, and Graft Loss in Deceased Donor Kidney Transplantation. Transplantation Direct 2019 Jan;5(1):e414 [FREE Full text] [doi: 10.1097/txd.0000000000000856]
10. Rao P, Schaubel D, Guidinger MK, Andreoni KA, Wolfe RA, Merion RM, et al. A comprehensive risk quantification score for deceased donor kidneys: the kidney donor risk index. Transplantation 2009 Jul 27;88(2):231-236. [doi: 10.1097/TP.0b013e3181ac620b] [Medline: 19623019]
11. Young A, Knoll GA, McArthur E, Dixon SN, Garg AX, Lok CE, et al. Is the Kidney Donor Risk Index a Useful Tool in Non-US Patients? Can J Kidney Health Dis 2018;5:2054358118791148 [FREE Full text] [doi: 10.1177/2054358118791148] [Medline: 30083367]
12. Senanayake S, White N, Graves N, Healy H, Baboolal K, Kularatna S. Machine learning in predicting graft failure following kidney transplantation: A systematic review of published predictive models. Int J Med Inform 2019 Oct;130:103957 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.103957] [Medline: 31472443]
13. Rao V, Behara R, Agrawal A. Predictive modeling for organ transplantation outcomes. 2014 Nov 10 Presented at: Proc - IEEE 14th Int Conf Bioinforma Bioeng BIBE 405?408; 2014; Boca Raton, FL, USA URL: https://ieeexplore.ieee.org/document/7033613/ [doi: 10.1109/BIBE.2014.58]
14. Brown T, Elster E, Stevens K, Graybill JC, Gillern S, Phinney S, et al. Bayesian modeling of pretransplant variables accurately predicts kidney graft survival. Am J Nephrol 2012;36(6):561-569 [FREE Full text] [doi: 10.1159/000345552] [Medline: 23221105]

15. Topuz K, Zengul F, Dag A, Almehmi A, Yildirim M. Predicting graft survival among kidney transplant recipients: A Bayesian decision support model. Decision Support Systems 2018 Feb;106:97-109 [FREE Full text] [doi: 10.1016/j.dss.2017.12.004]
16. Lin R, Horn S, Hurdle J, Goldfarb-Rumyantzev AS. Single and multiple time-point prediction models in kidney transplant outcomes. J Biomed Inform 2008 Dec;41(6):944-952 [FREE Full text] [doi: 10.1016/j.jbi.2008.03.005] [Medline: 18442951]
17. Axelrod D, Schnitzler M, Xiao H, Irish W, Tuttle-Newhall E, Chang SH, et al. An economic assessment of contemporary kidney transplant practice. Am J Transplant 2018 May;18(5):1168-1176 [FREE Full text] [doi: 10.1111/ajt.14702] [Medline: 29451350]
18. Lee D, Kanellis J, Mulley WR. Allocation of deceased donor kidneys: A review of international practices. Nephrology (Carlton) 2019 Jun;24(6):591-598 [FREE Full text] [doi: 10.1111/nep.13548] [Medline: 30536674]
19. George B, Seals S, Aban I. Survival analysis and regression models. J Nucl Cardiol 2014 Aug;21(4):686-694 [FREE Full text] [doi: 10.1007/s12350-014-9908-2] [Medline: 24810431]
20. Li J, Serpen G, Selman S. Bayes net classifiers for prediction of renal graft status and survival period. World Academy of Science, Engineering and Technology, Open Science Index 39, International Journal of Biomedical and Biological Engineering 2010:88-94. [doi: 10.5281/zenodo.1334844]
21. Mamoshina P, Vieira A, Putin E, Zhavoronkov A. Applications of Deep Learning in Biomedicine. Mol Pharm 2016 May 02;13(5):1445-1454 [FREE Full text] [doi: 10.1021/acs.molpharmaceut.5b00982] [Medline: 27007977]
22. Sadati N, Zafar M, Babu R. Representation Learning with Autoencoders for Electronic Health Records: A Comparative Study. Computing Research Repository. URL: https://arxiv.org/abs/1908.09174 [accessed 2020-11-20]
23. Dag A, Oztekin A, Yucel A, Bulur S, Megahed F. Predicting heart transplantation outcomes through data analytics. Decision Support Systems 2017 Feb;94:42-52 [FREE Full text] [doi: 10.1016/j.dss.2016.10.005]
24. Yoon J, Zame W, Banerjee A, Cadeiras M, Alaa AM, van der Schaar M. Personalized survival predictions via Trees of Predictors: An application to cardiac transplantation. PLoS One 2018;13(3):e0194985 [FREE Full text] [doi: 10.1371/journal.pone.0194985] [Medline: 29590219]
25. Van Belle V, Pelckmans K, Van Huffel S, Suykens JAK. Support vector methods for survival analysis: a comparison between ranking and regression approaches. Artif Intell Med 2011 Oct;53(2):107-118 [FREE Full text] [doi: 10.1016/j.artmed.2011.06.006] [Medline: 21821401]
26. Dorado-Moreno M, Pérez-Ortiz M, Gutiérrez PA, Ciria R, Briceño J, Hervás-Martínez C. Dynamically weighted evolutionary ordinal neural network for solving an imbalanced liver transplantation problem. Artif Intell Med 2017 Mar;77:1-11. [doi: 10.1016/j.artmed.2017.02.004] [Medline: 28545607]
27. Hong K, Iribarne A, Worku B, Takayama H, Gelijns AC, Naka Y, et al. Who is the high-risk recipient? Predicting mortality after heart transplant using pretransplant donor and recipient risk factors. Ann Thorac Surg 2011 Aug;92(2):520-7; discussion 527 [FREE Full text] [doi: 10.1016/j.athoracsur.2011.02.086] [Medline: 21683337]
28. Lofaro D, Maestripieri S, Greco R, Papalia T, Mancuso D, Conforti D, et al. Prediction of chronic allograft nephropathy using classification trees. Transplant Proc 2010 May;42(4):1130-1133 [FREE Full text] [doi: 10.1016/j.transproceed.2010.03.062] [Medline: 20534242]
29. Medved D, Nugues P, Nilsson J. Predicting the outcome for patients in a heart transplantation queue using deep learning. Annu Int Conf IEEE Eng Med Biol Soc 2017 Jul;2017:74-77. [doi: 10.1109/EMBC.2017.8036766] [Medline: 29059814]
30. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. jair 2002 Jun 01;16:321-357 [FREE Full text] [doi: 10.1613/jair.953]
31. Varoquaux G, Buitinck L, Louppe G, Grisel O, Pedregosa F, Mueller A. Scikit-learn. GetMobile: Mobile Comp. and Comm 2015 Jun;19(1):29-33 [FREE Full text] [doi: 10.1145/2786984.2786995]
32. Chollet F. Keras. URL: https://github.com/fchollet/keras [accessed 2020-09-15]
33. Medved D, Nugues P, Nilsson J. Selection of an optimal feature set to predict heart transplantation outcomes. Annu Int Conf IEEE Eng Med Biol Soc 2016 Aug;2016:3290-3293 [FREE Full text] [doi: 10.1109/EMBC.2016.7591431] [Medline: 28269008]
34. Thongkam J, Xu G, Zhang Y. AdaBoost algorithm with random forests for predicting breast cancer survivability. 2008 Presented at: IEEE International Joint Conference on Neural Networks; 1-8 June 2008; Hong Kong, China URL: https://doi.org/10.1109/IJCNN.2008.4634231 [doi: 10.1109/IJCNN.2008.4634231]
35. Tennankore K, Kim S, Alwayn I, Kiberd BA. Prolonged warm ischemia time is associated with graft failure and mortality after kidney transplantation. Kidney Int 2016 Mar;89(3):648-658 [FREE Full text] [doi: 10.1016/j.kint.2015.09.002] [Medline: 26880458]
36. Piros P, Ferenci T, Fleiner R, Andréka P, Fujita H, Főző L, et al. Comparing machine learning and regression models for mortality prediction based on the Hungarian Myocardial Infarction Registry. Knowledge-Based Systems 2019 Sep;179:1-7 [FREE Full text] [doi: 10.1016/j.knosys.2019.04.027]
37. Adegoke V, Chen D, Banissi E, Barikzai S. Prediction of breast cancer survivability using ensemble algorithms. 2017 Oct 18 Presented at: International Conference on Smart Systems and Technologies (SST); 2017; Osijek, Croatia URL: https://doi.org/10.1109/SST.2017.8188699 [doi: 10.1109/SST.2017.8188699]

38. Yoo K, Noh J, Lee H, Kim DK, Lim CS, Kim YH, et al. A Machine Learning Approach Using Survival Statistics to Predict Graft Survival in Kidney Transplant Recipients: A Multicenter Cohort Study. Sci Rep 2017 Aug 21;7(1):8904 [FREE Full text] [doi: 10.1038/s41598-017-08008-8] [Medline: 28827646]
39. Tiong H, Goldfarb D, Kattan MW, Alster JM, Thuita L, Yu C, et al. Nomograms for predicting graft function and survival in living donor kidney transplantation based on the UNOS Registry. J Urol 2009 Mar;181(3):1248-1255 [FREE Full text] [doi: 10.1016/j.juro.2008.10.164] [Medline: 19167732]
40. Luck M, Sylvain T, Cardinal H. Deep Learning for Patient-Specific Kidney Graft Survival Analysis. ArXiv. 2017. URL: https://arxiv.org/abs/1705.10245 [accessed 2020-10-24]
41. Loupy A, Aubert O, Orandi BJ, Naesens M, Bouatou Y, Raynaud M, et al. Prediction system for risk of allograft loss in patients receiving kidney transplants: international derivation and validation study. BMJ 2019 Sep 17;366:l4923 [FREE Full text] [doi: 10.1136/bmj.l4923] [Medline: 31530561]

Abbreviations

AdaBoost: adaptive boosting
ANN: artificial neural network
AUC: area under the curve
CMV: cytomegalovirus
ESRDDXSIMP: end-stage renal disease diagnosis
LR: logistic regression
ML: machine learning
RF: random forest
SMOTE: Synthetic Minority Oversampling Technique
SRTR: Scientific Registry of Transplant Recipients
SVM: support vector machine
VINTAGE: number of years on dialysis before transplant

Edited by R Kukafka; submitted 04.01.21; peer-reviewed by P Giabbanelli, E Mahmoudi; comments to author 01.02.21; revised version received 10.03.21; accepted 06.05.21; published 27.08.21.

Please cite as:
Naqvi SAA, Tennankore K, Vinson A, Roy PC, Abidi SSR
Predicting Kidney Graft Survival Using Machine Learning Methods: Prediction Model Development and Feature Significance Analysis Study
J Med Internet Res 2021;23(8):e26843
URL: https://www.jmir.org/2021/8/e26843
doi:10.2196/26843
PMID:34448704

©Syed Asil Ali Naqvi, Karthik Tennankore, Amanda Vinson, Patrice C Roy, Syed Sibte Raza Abidi. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Real-Time Respiratory Tumor Motion Prediction Based on a Temporal Convolutional Neural Network: Prediction Model Development Study

Panchun Chang[1,2*], MA; Jun Dang[1*], PhD; Jianrong Dai[3], PhD; Wenzheng Sun[4], PhD

[1]Department of Oncology, The First Affiliated Hospital of Chongqing Medical University, Chongqing, China

[2]School of Physics and Electronics, Shandong Normal University, Jinan, China

[3]Department of Radiation Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

[4]Department of Radiation Oncology, School of Medicine, The Second Affiliated Hospital, Zhejiang University, Hangzhou, China

[*]these authors contributed equally

Corresponding Author:
Wenzheng Sun, PhD
Department of Radiation Oncology, School of Medicine
The Second Affiliated Hospital
Zhejiang University
88 Jiefang Road
Hangzhou, 310009
China
Phone: 86 057187783538
Email: sunwenzheng@zju.edu.cn

Abstract

Background: The dynamic tracking of tumors with radiation beams in radiation therapy requires the prediction of real-time target locations prior to beam delivery, as treatment involving radiation beams and gating tracking results in time latency.

Objective: In this study, a deep learning model that was based on a temporal convolutional neural network was developed to predict internal target locations by using multiple external markers.

Methods: Respiratory signals from 69 treatment fractions of 21 patients with cancer who were treated with the CyberKnife Synchrony device (Accuray Incorporated) were used to train and test the model. The reported model's performance was evaluated by comparing the model to a long short-term memory model in terms of the root mean square errors (RMSEs) of real and predicted respiratory signals. The effect of the number of external markers was also investigated.

Results: The average RMSEs of predicted (ahead time=400 ms) respiratory motion in the superior-inferior, anterior-posterior, and left-right directions and in 3D space were 0.49 mm, 0.28 mm, 0.25 mm, and 0.67 mm, respectively.

Conclusions: The experiment results demonstrated that the temporal convolutional neural network–based respiratory prediction model could predict respiratory signals with submillimeter accuracy.

(J Med Internet Res 2021;23(8):e27235) doi:10.2196/27235

KEYWORDS

radiation therapy; temporal convolutional neural network; respiratory signal prediction; neural network; deep learning model; dynamic tracking

Introduction

The aim of radiation therapy is not only to deliver lethal doses of radiation to target tumors but also to minimize the dose of unnecessary radiation delivered to the surrounding healthy tissues and structures [1-5]. Modern technical advances, such as intensity-modulated radiation therapy, have improved the accuracy of dose delivery. However, some targets, such as lung cancer and liver cancer tumors, may move substantially during the treatment delivery process due to respiratory motion [6-10]. Investigators have reported that lung and liver tumors can move up to 3 cm during a conventional radiation therapy treatment session [11,12]. The motion of targets may substantially decrease the accuracy and efficiency of intensity-modulated radiation therapy or other advanced technologies.

Many methods have been investigated to reduce the effect of respiratory motion, which mainly include the following:

- Adding a margin around the target tumor: a 10- to 15-mm margin is always used as the radiation treatment field to avoid missing a tumor, which may result in unnecessary radiation exposure to heathy tissues and structures [13].
- Breath hold: patients need to hold their breath during the treatment to temporarily stop respiration, but this is not applicable for some patients, such as older patients and juvenile patients [14].
- Beam tracking: radiation beams track a moving tumor dynamically to ensure that the tumor target is constantly within the treatment field [15].

All beam tracking methods must compensate for the latency of various sources, such as latencies from beam adjustment and image capture times [5,16]. Hence, we must estimate the position of targets in advance to compensate for latency effects.

Recently, deep learning approaches based on long short-term memory (LSTM) have been successfully used to solve time series prediction problems in several fields. For example, Ma et al [17] used an LSTM model to capture traffic dynamics data for predicting short-term traffic speed. Bao et al [18] implemented an LSTM model to predict the one-step-ahead price (closing) of 6 stock indices for various financial markets. Lin et al [19] used an LSTM model to predict respiratory signals. Moreover, some recent studies have demonstrated that certain temporal convolutional neural network (TCN) architectures could achieve state-of-the-art accuracy in time series prediction problems [20-23]. However, to our knowledge, there are no studies on using a TCN model to predict respiratory tumor motion. Hence, in this study, we developed a TCN-based respiratory prediction model by using external markers and compared the prediction performance of the TCN to that of an LSTM model. We also investigated the effect that the number of external markers had on prediction performance.

Methods

Data Acquisition

The tumor motion data (69 treatment fractions of 21 patients) used in this study were obtained from an open data set, which was recorded by the CyberKnife Synchrony (Accuray Incorporated) tracking system with a recorded sampling rate of 25 Hz [24]. To analyze the external movements of patients, charge-coupled device cameras were used to monitor the luminous diodes located on a patient's abdomen and chest. To analyze internal fiducial positions, orthogonal diagnostic x-ray systems were used to observe implanted markers periodically.

Prediction Process

The general scheme for the prediction process of 2 models is outlined in Figure 1, and the arrangement of the respiratory signals that were used for network training and validation is shown in Table 1. Each recorded position (internal tumor and external marker positions) was stratified into 2 cohorts based on time t_s. The positions prior to time t_s (the training signals) were used to train the TCN and LSTM models. The positions after t_s (the testing signals) were used to evaluate the developed model.

Figure 1. Flowchart of the prediction algorithm.

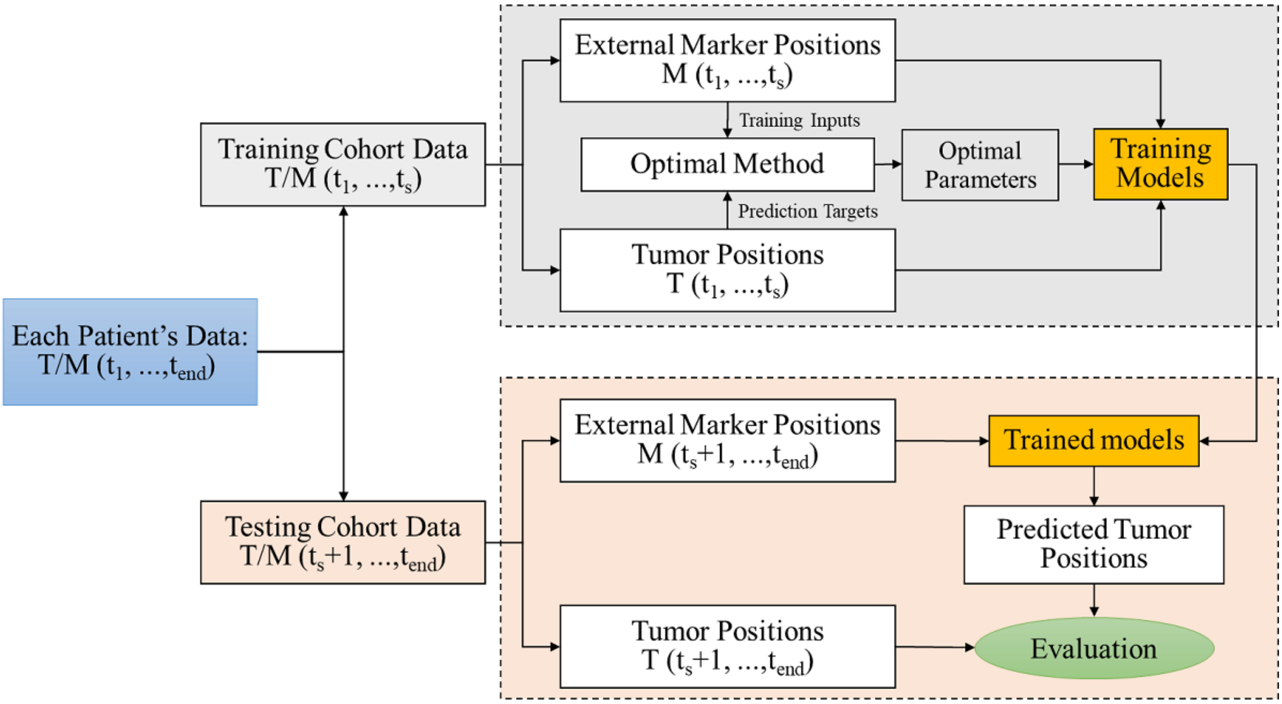


Table 1. The arrangement of respiratory signals used for network training and validation.

Position type	Data for training	Data for validation
Inputs of the network		
Position of marker 1	M[a]1_{SI}[b], $_{AP}$[c], $_{LR}$[d] (1, 2,…, t_s)	$M1_{SI, AP, LR}$ (t_s+1, 2,…, t_s+t_{end})
Position of marker 2	$M2_{SI, AP, LR}$ (1, 2,…, t_s)	$M2_{SI, AP, LR}$ (t_s+1, 2,…, t_s+t_{end})
Position of marker 3	$M3_{SI, AP, LR}$ (1, 2,…, t_s)	$M3_{SI, AP, LR}$ (t_s+1, 2,…, t_s+t_{end})
Targets of the network		
Position of a tumor	T[e]$_{SI, AP, LR}$ (1, 2,…, t_s)	$T_{SI, AP, LR}$ (t_s+1, 2,…, t_s+t_{end})

[a] M: external marker position.

[b] SI: superior-inferior.

[c] AP: anterior-posterior.

[d] LR: left-right.

[e] T: tumor position.

For the training process, the training input data and prediction target data were first used to tune the hyperparameters, which was done by using a cross-validation model. Afterward, they were used to train the model. The external markers' positions during the first input period of the training process (ie, the time between t=1 and $t=t_{delay}$) were used as the training input data for predicting the tumor positions (target positions) at a specific time frame ($t=t_{delay}+t_{ahead}$). This training process was repeated and continued to predict the next tumor position until either the threshold of the cost function or the maximum iteration number, which was set in advance, was reached. Each pair of data points (ie, the input data, M[t+1,…, $t+t_{delay}$], vs the output data, T[$t+t_{delay}+t_{ahead}$]) consisted of a training data set. "M" denoted 3 external markers' positions (M1, M2, and M3), which were based on 3 directions (the superior-inferior, anterior-posterior, and left-right directions). t_{ahead} represented the ahead time we needed for making predictions.

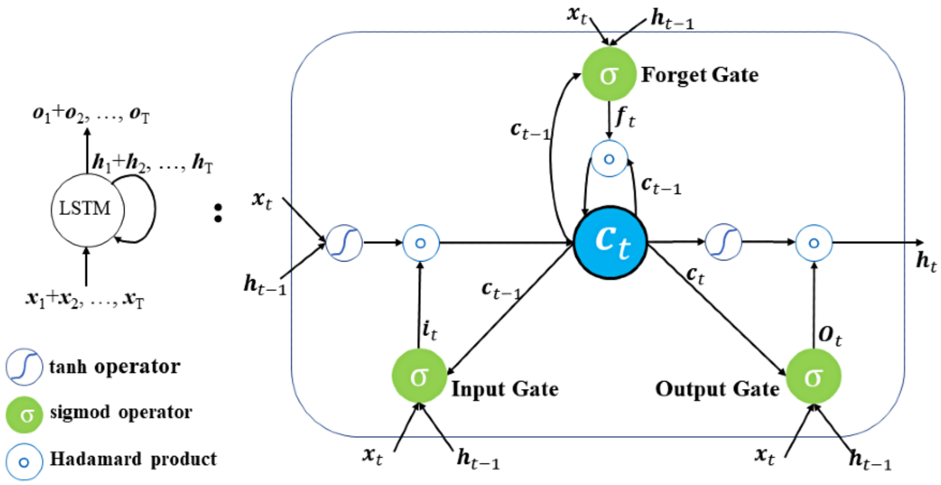
For the evaluation process, the testing signals, which were represented as M(t_s+1, t_s+2,…, t_{end}) and T(t_s+1, t_s+2,…, t_s+ t_{end}), were used to evaluate the developed model. Similar to the process implemented in the training process, the external markers' positions during the first input period of the evaluation process (ie, the time between t=1 and $t=t_{delay}$) were used to predict a tumor's position (T'[$t_s+t_{delay}+t_{ahead}$]) at a specific time ($t=t_s+t_{delay}+t_{ahead}$). This process was also repeated to predict the next tumor position continuously. The external signals that were recorded during radiation therapy (ie, the time between $t=t_{end}-t_{delay}-t_{ahead}+1$ and $t=t_{end}-t_{ahead}$) were used to predict the final tumor position (T'[t_{end}]). Finally, the predicted signals (T'[$t_s+t_{delay}+t_{ahead}$],…, T'[t_{end}]) were compared to the real tumor positions (T[$t_s+t_{delay}+t_{ahead}$],…, T[t_{end}]).

LSTM Model

The recurrent neural network (RNN) is a particular type of neural network that allows for self-cycle connections and transmits parameters across different time stamps. An RNN model can store the information of former time stamps. However, it is difficult for the RNN to memorize long-term memory information due to vanishing and exploding gradients [25-27].

The LSTM layer is a special RNN layer that overcomes the weakness that the RNN has with memorizing long-term memory information [26,28]. Figure 2 shows an LSTM unit. Unlike the simple RNN unit, the LSTM unit has a memory cell state c_t at time t. The information that passes through state c_t is controlled by the following three gates: the input gate (i_t), the forget gate (f_t), and the output gate (o_t). The input gate is used to control input data that flow into state c_t, the hidden state connection (h_t) is used to control the forgetting of state c_t, and the output gate is used to moderate the output data that flow from state c_t. A plurality of LSTM layers can be stacked in a deeper neural network, which can fit the data of the complicated functions that are required to analyze the inputs and the targets.

Figure 2. The structure of an LSTM layer. LSTM: long short-term memory.


$o_1+o_2, \dots, o_T$
$h_1+h_2, \dots, h_T$
LSTM
$x_1+x_2, \dots, x_T$
:
tanh operator
σ sigmod operator
Hadamard product
x_t
h_{t-1}
Forget Gate
f_t
c_{t-1}
c_{t-1}
C_t
c_{t-1}
c_t
i_t
o_t
h_t
Input Gate
Output Gate
x_t
h_{t-1}
x_t
h_{t-1}

TCN Model

The TCN model was based on a transformation of a 1D fully convolutional network that was used for sequential prediction problems. The TCN model used a multilayer network to learn information over a long time span. Sequence information were transmitted layer by layer across the network until prediction results were obtained. The architecture of the TCN model is illustrated in Figure 3 [23], in which $x_1, x_2, \dots, x_T$ are the original sequence signals (inputs), and are the prediction signals (outputs). The obvious characteristics of the TCN model, which were compared to those of the normal 1D fully convolutional network model, were as follows:

- The TCN model used causal convolutions, in which the output at time t was convolved only with elements from previous layers at time t and earlier, to ensure that no leakage occurred from the future into the past.
- The TCN model used dilated convolutions to ensure that each hidden layer had the same size as the input sequence and to increase the receptive field (ie, learning longer lengths of information).

The input of the TCN model was interval sampled. The equation for the dilated convolution was as follows:

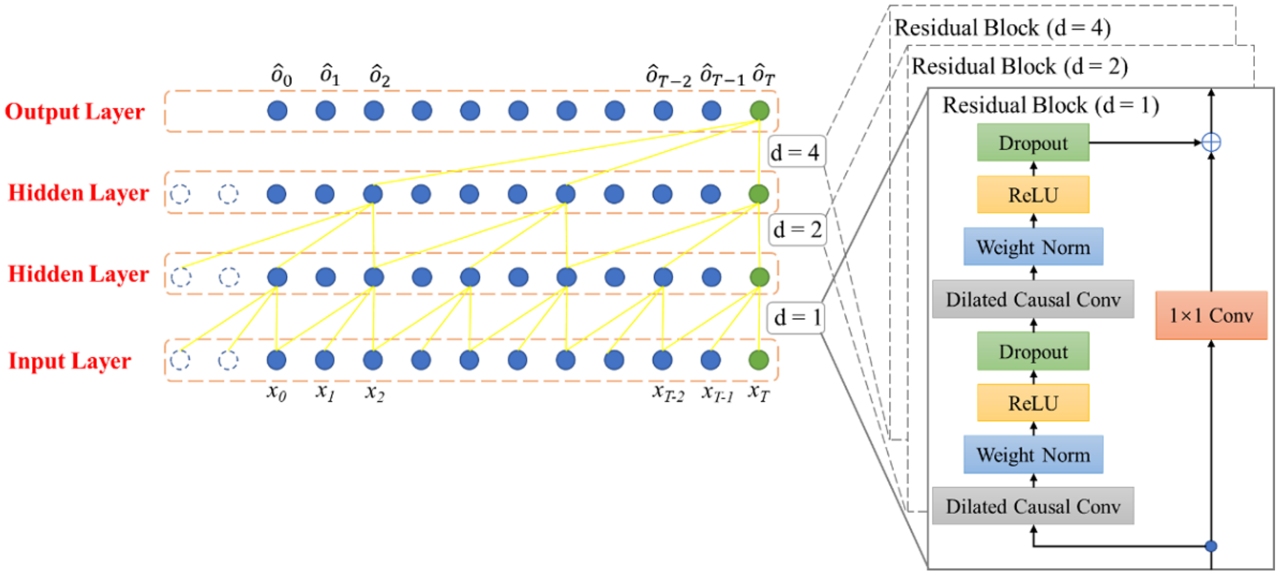
In equation 1, *d* is the dilation factor (sampling rate). A *d* value of 1 in the lowest layer meant that every signal was sampled, whereas a *d* value of 2 in the middle layer meant that every 2 respiratory signals were sampled.

Residual networks [29], which are shown in Figure 3, were imported in this study to accelerate convergence and stabilize training. A residual block that included a branch was used to make a series of transformations (*F*). Afterward, the outputs of the residual block (ie, $F[X_{residual}]$) were added to the input (ie, $X_{residual}$), as follows:

$$O_{residual} = \text{Activation}(X_{residual} + F[X_{residual}])$$

(2)

Figure 3. The architecture of the temporal convolutional neural network model. "d" was the dilation factor. Conv: convolution; ReLU: rectified linear unit.


Output Layer
Hidden Layer
Hidden Layer
Input Layer
$\hat{o}_0$ $\hat{o}_1$ $\hat{o}_2$ $\hat{o}_{T-2}$ $\hat{o}_{T-1}$ $\hat{o}_T$
x_0 x_1 x_2 x_{T-2} x_{T-1} x_T
d = 4
d = 2
d = 1
Residual Block (d = 4)
Residual Block (d = 2)
Residual Block (d = 1)
Dropout
ReLU
Weight Norm
Dilated Causal Conv
Dropout
ReLU
Weight Norm
Dilated Causal Conv
1×1 Conv

Hyperparameter Tuning

With regard to the TCN model, previous TCN studies [20-23] reported (in the *Instruction* section) using the same TCN architecture and only sometimes varying the number of layers (n) and the filter size. Hence, we tested these two hyperparameters and used a dilation factor (d) of 2^n for layer n. Moreover, the number of neurons in the input layer and the learning rate of the TCN model were also investigated in this study. For the LSTM model, the number of LSTM layers, learning rate, number of hidden units per layer, and number of neurons in the input layer were investigated. Furthermore, the Adam algorithm was used as the optimization algorithm for both the TCN model and LSTM model. The Kingma and Ba [30] study demonstrated that the hyperparameters in the Adam model required little tuning. Goodfellow et al [31] also approved of the robustness of the Adam model for their hyperparameter of choice and provided advice on how to tune the learning rate from the default value. Hence, we used the good default settings that were tested by Kingma and Ba [30] as the hyperparameters of the Adam optimizer and tuned the learning rate. The default settings were exponential decay rates of 0.9 and 0.999 and a decay exponent of 10^{-8}. In this study, all hyperparameters were tuned synthetically by using a grid search model. It should be noted that we tested the hyperparameters in a 4D hyperparameter space instead of a subspace (ie, while a parameter was investigated, others were fixed) to maintain the accuracy of hyperparameter tuning.

Model Evaluation

The respiratory signals from 69 treatment fractions of 21 patients with cancer who were treated with the CyberKnife Synchrony (Accuray Incorporated) device were used to evaluate the proposed model. Of the 69 treatment fractions, 5 were used to tune the hyperparameters. The rest of the patients were used to evaluate prediction performance. For each of the 69 treatment fractions, signals that were acquired around the first 3 minutes (4500 data points) were used as the training signals for training the prediction model, and signals from the following 30 seconds were used as the test signals for assessing the effectiveness of the proposed model. The ahead time (t_{ahead}) used in this study was 400 ms [1,5].

The root mean square errors (RMSEs) between real and predicted signals of respiratory motion in a 3D space were used for assessment [6,7]. The RMSEs for motion in each direction ($RMSE_{SI,\ LR,\ AP}$) and motion in a 3D space ($RMSE_{3D}$) were calculated by using equations 3 and 4, respectively, as follows:

In equation 5, is the average of the true values, and is the average of predicted values. Time point t in equation 3 ranged from t_{start} (t_s+t_{delay}+t_{ahead}) to t_{end}. The Wilcoxon signed-rank test was used as the statistical model for evaluating the differences between true values and predicted values.

Results

Table 2 presents the RMSEs of the three models (ie, the LSTM, TCN, and no prediction models; ahead time=400 ms). Compared to the no prediction model, the RMSEs for motion in a 3D space were reduced by 46% in the LSTM model and 51% in the TCN model. For motion in all directions, the RMSEs of the TCN model were consistently lower than those of the LSTM model. The RMSE for motion in a 3D space decreased from 0.73 mm (LSTM model) to 0.67 mm (TCN model). The P value was <.001, indicating that the TCN method could significantly improve the prediction performance of the LSTM method.

Table 2. The root mean square errors (RMSEs) of the three prediction models.

Direction	RMSEs (mm) of the LSTM[a] model	RMSEs (mm) of the TCN[b] model	RMSEs (mm) of the no prediction model
Anterior-posterior direction	0.29	0.28	0.50
Left-right direction	0.27	0.25	0.45
Superior-inferior direction	0.55	0.49	1.04
3D space	0.73	0.67	1.36

[a] LSTM: long short-term memory.

[b] TCN: temporal convolutional neural network.

Figure 4 shows the RMSEs for motion in all directions with different ahead times. Obviously, the prediction performance of the TCN model was positive compared to that of the LSTM model for all ahead times. Further, the prediction performance of both models worsened as ahead times increased.

Figure 5 illustrates the performance comparison between the TCN and LSTM methods for predicting motion in the superior-inferior direction, anterior-posterior direction, and left-right direction. Obviously, the TCN method was more accurate and robust than the LSTM method.

We investigated the hyperparameters in the 4D hyperparameter space (625 experiments) for both the TCN and LSTM models by using the grid search method among 5 treatment fractions, which were selected randomly. The options and results of hyperparameter tuning are depicted in Table 3.

Figure 4. The RMSEs for respiratory motion in all directions. These were determined by using the LSTM and TCN models and different ahead times for each treatment fraction. AP: anterior-posterior; LR: left-right; LSTM: long short-term memory; RMSE: root mean square error; SI: superior-inferior; TCN: temporal convolutional neural network.

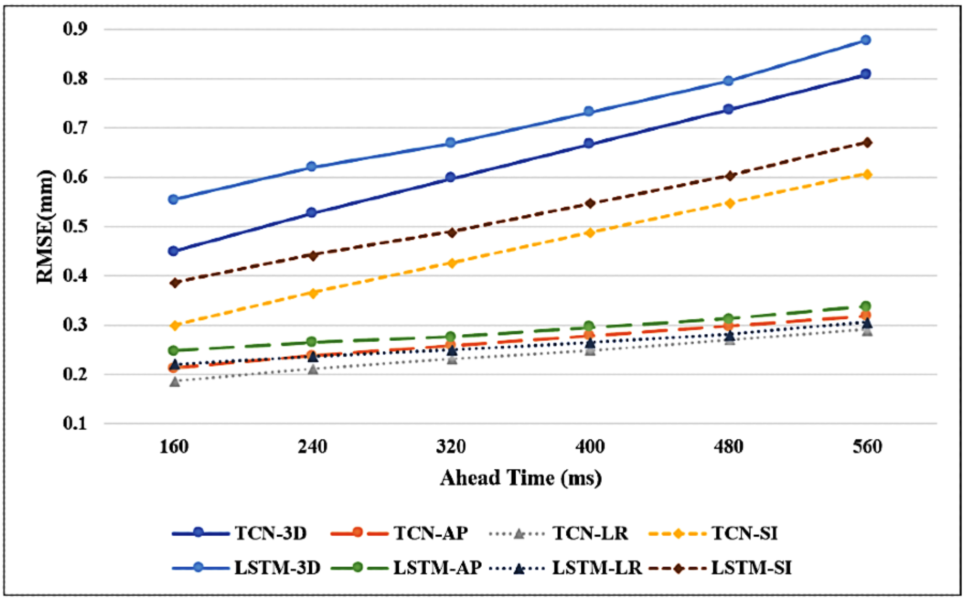


Figure 5. The performance comparison between the TCN and LSTM methods for predicting motion in the (A) superior-inferior direction, (B) left-right direction, and (C) anterior-posterior direction. LSTM: long short-term memory; TCN: temporal convolutional neural network.

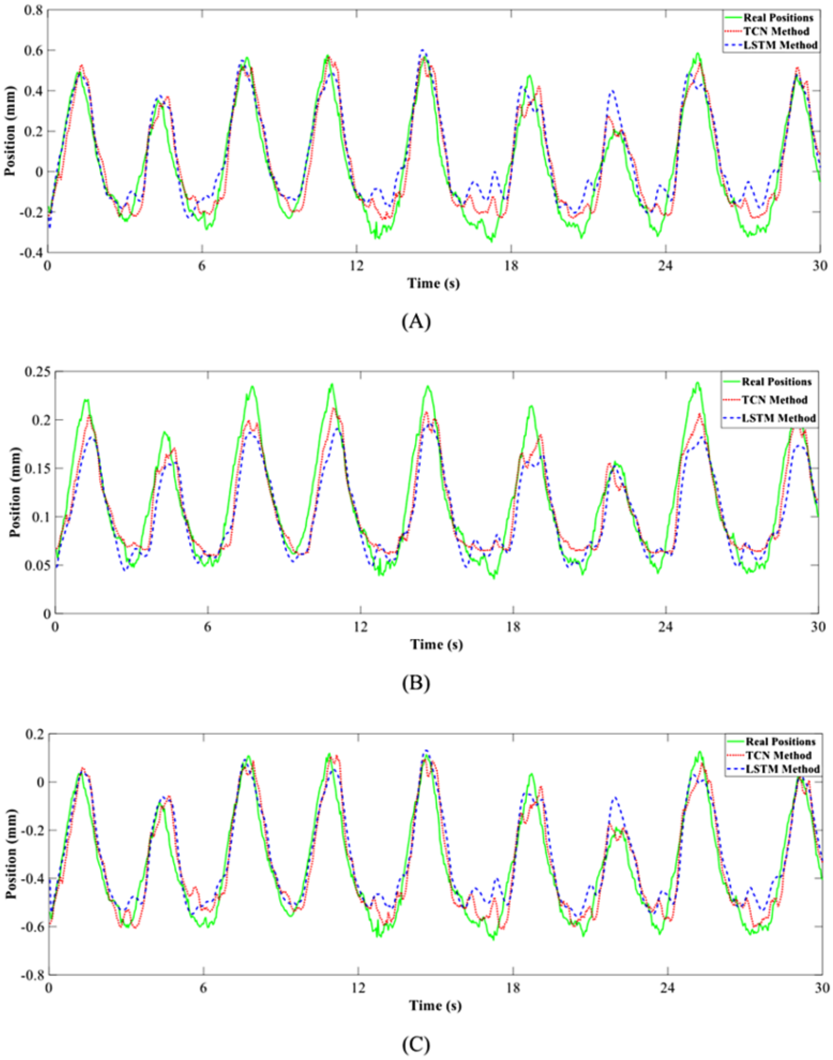


Table 3. The options and results of hyperparameter tuning.

Models and hyperparameters	Hyperparameter options	Hyperparameter selected
Temporal convolutional neural network model		
Number of layers	4, 5, 6, 7, and 8	5
Filter size	1, 3, 5, 7, and 9	9
Number of neurons in the input layer	5, 10, 15, 20, and 25	15
Learning rate	0.0001, 0.001, 0.005, 0.01, and 0.1	0.001
LSTM[a] model		
Number of LSTM layers	1, 2, 3, 4, and 5	2
Learning rate	0.0001, 0.001, 0.005, 0.01, and 0.1	0.01
Number of hidden units per layer	10, 50, 100, 150, 200, and 250	200
Number of neurons in the input layer	5, 10, 15, 20, and 25	20

[a] LSTM: long short-term memory.

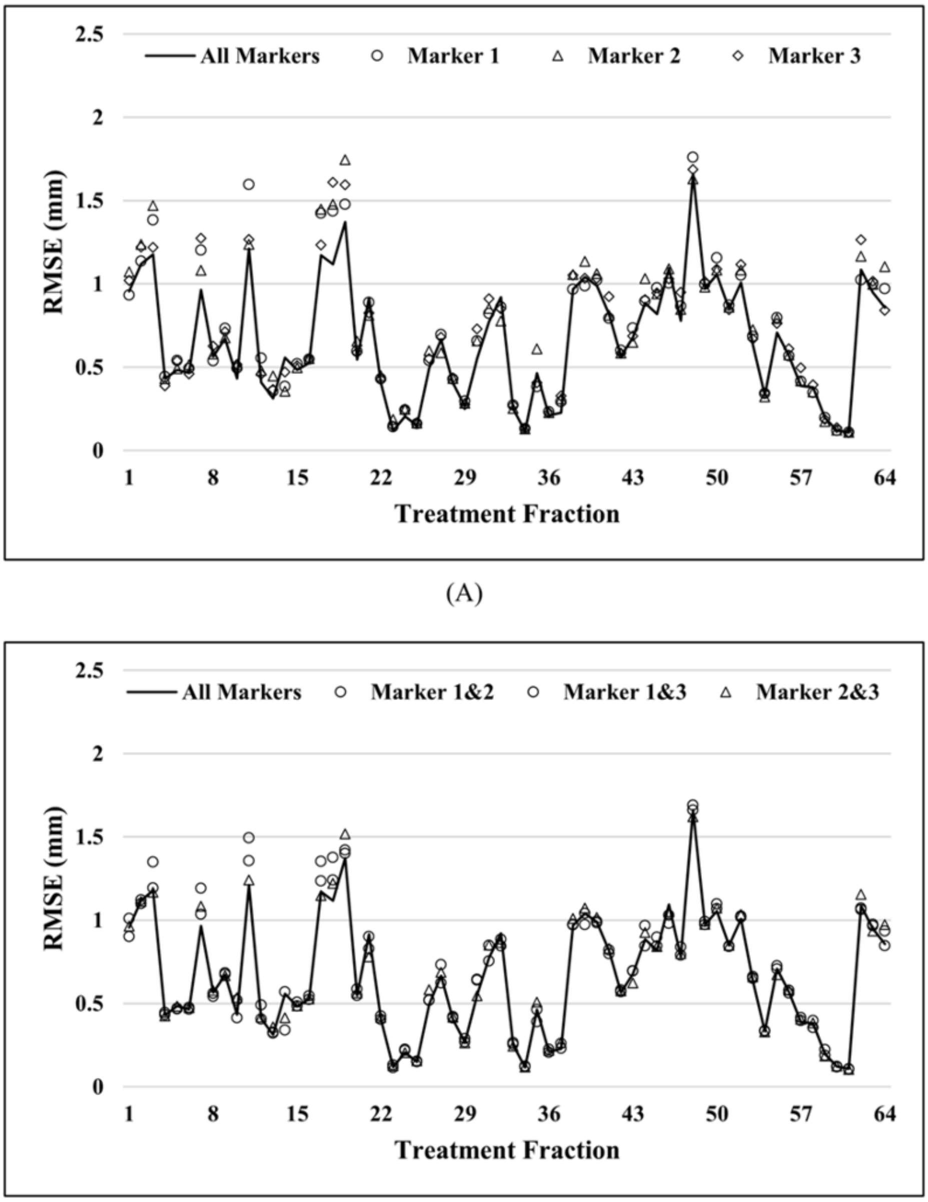
Table 4 presents the RMSEs of the TCN model for each external marker. Figure 6 shows the RMSEs for respiratory motion in a 3D space among each treatment fraction. The TCN model using 1 or 2 external markers was compared to the TCN model using all 3 external markers. The TCN model had the best performance in terms of predicting motion in all directions when all three external markers were used simultaneously. The average RMSEs for motion in a 3D space when using 1 marker and 2 markers were 0.72 mm and 0.68 mm, respectively. This decreased to 0.67 mm when using all three makers.

As illustrated in Figure 7, the ablative analysis of the TCN was also conducted. We focused on two components in this study—the filter size and the residual blocks. We found that the effect of the filter size was small when the filter size was larger than 3. The *P* values between 5 filter size pairs—filter sizes 1 and 3, 3 and 5, 5 and 7, and 7 and 9—were <.001, .11, .20, and .83, respectively. This indicated that prediction performance improved significantly before the filter size rose to 3. Further, we found that the residual blocks contributed significantly to prediction performance, as the *P* value was <.001.

Table 4. The root mean square errors (RMSEs) of the temporal convolutional neural network model for each external marker (EM).

Direction	RMSEs for all EMs	RMSEs for EMs 1 and 2	RMSEs for EMs 1 and 3	RMSEs for EMs 2 and 3	RMSEs for EM 1	RMSEs for EM 2	RMSEs for EM 3
Anterior-posterior direction	0.28	0.28	0.28	0.28	0.29	0.29	0.29
Left-right direction	0.25	0.26	0.26	0.25	0.27	0.26	0.26
Superior-inferior direction	0.49	0.51	0.50	0.50	0.52	0.53	0.53
3D space	0.67	0.69	0.68	0.68	0.71	0.72	0.72

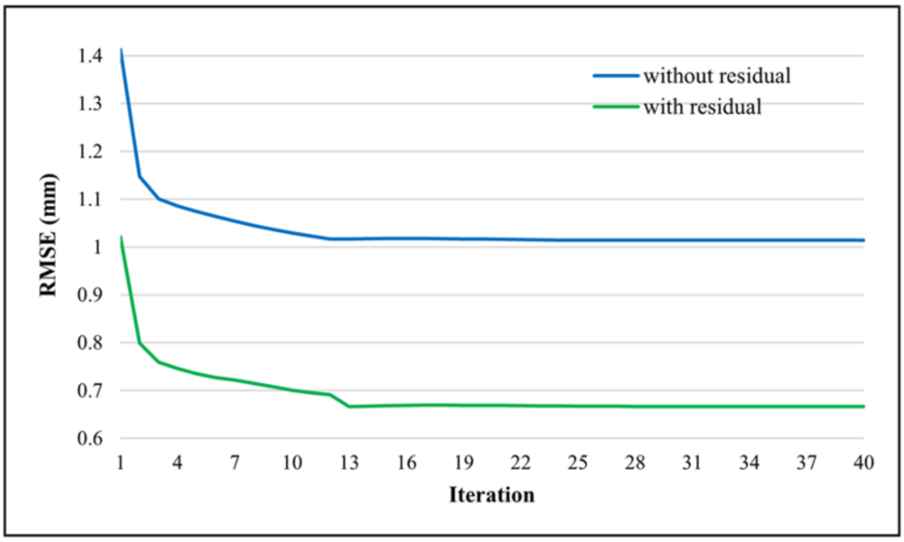
Figure 6. A comparison of RMSEs for respiratory motion in a 3D space among each treatment fraction. A: Results of the TCN model using 1 external marker compared to those of the TCN model using all 3 external markers. B: Results of the TCN model using 2 external markers compared to those of the TCN model using all 3 external markers. RMSE: root mean square error; TCN: temporal convolutional neural network.



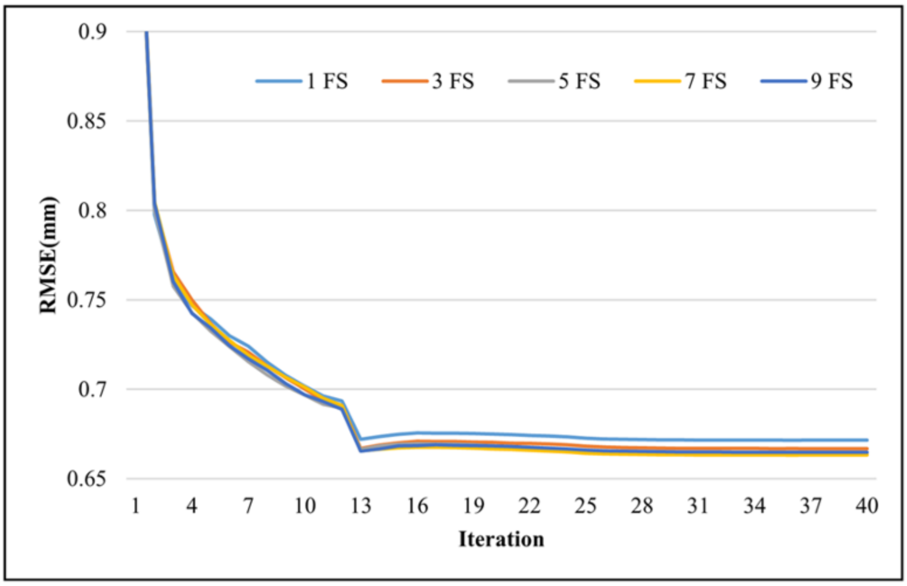
(A)

(B)

Figure 7. The effects of different components in the temporal convolutional neural network layer. A: Residual blocks. B: FS. FS: filter size; RMSE: root mean square error.



(A)



(B)

Discussion

Principal Findings

A TCN model for predicting respiratory motion by using external markers' prior signals was developed and tested in this study. The experiment demonstrated that the TCN model's performance in predicting future respiratory signals with a 400-ms ahead time was better than that of the LSTM model.

As is well known, hyperparameter settings have a large influence on the prediction performance of machine learning models. This also holds true for our TCN and LSTM models. We tuned 4 major hyperparameters for both of the TCN and LSTM models. Among these hyperparameters, the number of neurons in the input layer and the learning rate were tested for both models. Having a large number of neurons in the input layer allows for the inclusion of more features in models. Obviously, useful features may increase prediction accuracy. However, redundancy features may also be brought in along with the useful features. Hence, if this hyperparameter is too large, prediction performance may degenerate. The best number of neurons in the input layer for the TCN and LSTM models in this study was 15 and 20, respectively. The learning rate was an important hyperparameter in the model optimization process. If the learning rate is too large, the model may oscillate around the global minimum value instead of achieving convergence. On the other hand, if this value is too small, the training time and the risk of overfitting increase. Learning rates of 0.001 and 0.01 were selected as the optimal hyperparameters of the TCN and LSTM models, respectively. In addition to the two abovementioned hyperparameters, the number of layers and filter sizes were also investigated for the TCN model, whereas the number of LSTM layers and number of hidden units per layer were tested for the LSTM model. With regard to the TCN model, the size of the effective window (receptive field) increased as the number of layers and filter size increased. Hence, these two hyperparameters should guarantee that the receptive field of TCN model covers enough context for respiratory signal prediction. The optimal values for these two hyperparameters in our experiments were 5 and 9, respectively. With regard to the LSTM model, on one hand, a deeper LSTM model (a large number of LSTM layers) may be representative of a more complicated relationship among respiratory signals and improve prediction performance. On the other hand, a deeper LSTM model also has an increased risk of overfitting and increased convergence speed. In this study, the prediction performance results of the LSTM model were comparable when the number of LSTM layers was over 2. Hence, we selected 2 as the optimal number of LSTM layers. Further, the number of hidden units per layer determined the width of each LSTM layer. We also found that having a large number of hidden units per layer was helpful for establishing a more complicated prediction model, but at the same time, this increased the risk of overfitting and convergence speed.

The effect that different numbers of external markers had on prediction performance was also investigated in this study. The TCN model had the best prediction performance when it used all three markers' positions. As shown in Figure 6, the TCN model's prediction performance when using 3 markers was more robust than when using 1 marker or 2 markers. For most treatment fractions, the RMSEs of the TCN model using 3 markers was slightly smaller than those obtained by using 1 marker or 2 markers. However, for some treatment fractions, such as treatment fractions 7 and 11, the RMSEs of predictions based on 1 or 2 external markers were quite larger than those of predictions based on 3 external markers. This was probably because having more external markers for different skin surface positions resulted in the inclusion of more useful features. Such useful features may alleviate the overfitting and underfitting problems.

Finally, we studied the influence of the different components (the filter size and residual blocks) in the TCN model. The size of the effective window (receptive field) increased with filter size. Hence, the model's prediction performance initially became better as the filter size increased. However, the model's prediction performance only slightly improved as the filter size increased continually. This may be because the receptive field that resulted from using a filter size of 3 provided enough context for the respiratory signal prediction task. On the other hand, we observed that the residual block architecture enhanced the model's prediction performance immensely. We believe that this was because the residual blocks effectively allowed the TCN model to be modified based on identity mapping instead of a full transformation, which was crucial for the deep neural network architecture.

Conclusion

A deep learning approach based on the TCN architecture was developed to predict internal tumor positions with a 400-ms ahead time based on the external markers' positions in this study. The results demonstrated that this model could predict tumor positions accurately. Further, the prediction performance of the TCN model using multiple external markers was more robust and positive than that of the TCN model using 1 or 2 external markers.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (62103366), the General Project of Chongqing Natural Science Foundation (grant cstc2020jcyj-msxm2928), Seed Grant of the First Affiliated Hospital of Chongqing Medical University (grant PYJJ2019-208), Chongqing Municipal Bureau of Human Resources and Social Security Fund (grant cx2018147), and Medical Research Key Project of Jiangsu Health Commission (grant ZDB 2020022).

Conflicts of Interest

None declared.

References

1. Ren Q, Nishioka S, Shirato H, Berbeco RI. Adaptive prediction of respiratory motion for motion compensation radiotherapy. Phys Med Biol 2007 Nov 21;52(22):6651-6661. [doi: 10.1088/0031-9155/52/22/007] [Medline: 17975289]
2. McCall KC, Jeraj R. Dual-component model of respiratory motion based on the periodic autoregressive moving average (periodic ARMA) method. Phys Med Biol 2007 Jun 21;52(12):3455-3466. [doi: 10.1088/0031-9155/52/12/009] [Medline: 17664554]
3. Bukhari W, Hong SM. Real-time prediction and gating of respiratory motion using an extended Kalman filter and Gaussian process regression. Phys Med Biol 2015 Jan 07;60(1):233-252. [doi: 10.1088/0031-9155/60/1/233] [Medline: 25489980]
4. Verma P, Wu H, Langer M, Das I, Sandison G. Survey: Real-time tumor motion prediction for image-guided radiation treatment. Comput Sci Eng 2011 Sep;13(5):24-35. [doi: 10.1109/mcse.2010.99]
5. Riaz N, Shanker P, Wiersma R, Gudmundsson O, Mao W, Widrow B, et al. Predicting respiratory tumor motion with multi-dimensional adaptive filters and support vector regression. Phys Med Biol 2009 Oct 07;54(19):5735-5748. [doi: 10.1088/0031-9155/54/19/005] [Medline: 19729711]
6. Yan H, Yin FF, Zhu GP, Ajlouni M, Kim JH. The correlation evaluation of a tumor tracking system using multiple external markers. Med Phys 2006 Nov;33(11):4073-4084. [doi: 10.1118/1.2358830] [Medline: 17153387]

7. Sun WZ, Jiang MY, Ren L, Dang J, You T, Yin FF. Respiratory signal prediction based on adaptive boosting and multi-layer perceptron neural network. Phys Med Biol 2017 Aug 03;62(17):6822-6835 [FREE Full text] [doi: 10.1088/1361-6560/aa7cd4] [Medline: 28665297]
8. Ernst F, Schlaefer A, Schweikard A. Predicting the outcome of respiratory motion prediction. Med Phys 2011 Oct;38(10):5569-5581. [doi: 10.1118/1.3633907] [Medline: 21992375]
9. Torshabi AE, Riboldi M, Fooladi AAI, Mosalla SMM, Baroni G. An adaptive fuzzy prediction model for real time tumor tracking in radiotherapy via external surrogates. J Appl Clin Med Phys 2013 Jan 07;14(1):4008 [FREE Full text] [doi: 10.1120/jacmp.v14i1.4008] [Medline: 23318386]
10. Vergalasova I, Cai J, Yin FF. A novel technique for markerless, self-sorted 4D-CBCT: feasibility study. Med Phys 2012 Mar;39(3):1442-1451 [FREE Full text] [doi: 10.1118/1.3685443] [Medline: 22380377]
11. Berbeco RI, Nishioka S, Shirato H, Jiang SB. Residual motion of lung tumors in end-of-inhale respiratory gated radiotherapy based on external surrogates. Med Phys 2006 Nov;33(11):4149-4156. [doi: 10.1118/1.2358197] [Medline: 17153393]
12. Shirato H, Suzuki K, Sharp GC, Fujita K, Onimaru R, Fujino M, et al. Speed and amplitude of lung tumor motion precisely detected in four-dimensional setup and in real-time tumor-tracking radiotherapy. Int J Radiat Oncol Biol Phys 2006 Mar 15;64(4):1229-1236. [doi: 10.1016/j.ijrobp.2005.11.016] [Medline: 16504762]
13. Goodband JH, Haas OCL, Mills JA. A comparison of neural network approaches for on-line prediction in IGRT. Med Phys 2008 Mar;35(3):1113-1122. [doi: 10.1118/1.2836416] [Medline: 18404946]
14. Nelson C, Starkschall G, Balter P, Fitzpatrick MJ, Antolak JA, Tolani N, et al. Respiration-correlated treatment delivery using feedback-guided breath hold: a technical study. Med Phys 2005 Jan;32(1):175-181. [doi: 10.1118/1.1836332] [Medline: 15719968]
15. Hansen R, Ravkilde T, Worm ES, Toftegaard J, Grau C, Macek K, et al. Electromagnetic guided couch and multileaf collimator tracking on a TrueBeam accelerator. Med Phys 2016 May;43(5):2387. [doi: 10.1118/1.4946815] [Medline: 27147350]
16. Sharp GC, Jiang SB, Shimizu S, Shirato H. Prediction of respiratory tumour motion for real-time image-guided radiotherapy. Phys Med Biol 2004 Feb 07;49(3):425-440. [doi: 10.1088/0031-9155/49/3/006] [Medline: 15012011]
17. Ma X, Tao Z, Wang Y, Yu H, Wang Y. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. Transp Res Part C Emerg Technol 2015 May;54:187-197. [doi: 10.1016/j.trc.2015.03.014]
18. Bao W, Yue J, Rao Y. A deep learning framework for financial time series using stacked autoencoders and long-short term memory. PLoS One 2017 Jul 14;12(7):e0180944. [doi: 10.1371/journal.pone.0180944] [Medline: 28708865]
19. Lin H, Shi C, Wang B, Chan MF, Tang X, Ji W. Towards real-time respiratory motion prediction based on long short-term memory neural networks. Phys Med Biol 2019 Apr 10;64(8):085010 [FREE Full text] [doi: 10.1088/1361-6560/ab13fa] [Medline: 30917344]
20. Dauphin YN, Fan A, Auli M, Grangier D. Language modeling with gated convolutional networks. 2017 Aug Presented at: The 34th International Conference on Machine Learning; August 6-11, 2017; Sydney, Australia.
21. Gehring J, Auli M, Grangier D, Yarats D, Dauphin YN. Convolutional sequence to sequence learning. 2017 Aug Presented at: The 34th International Conference on Machine Learning; August 6-11, 2017; Sydney, Australia.
22. Kalchbrenner N, Espeholt L, Simonyan K, van den Oord A, Graves A, Kavukcuoglu K. Neural machine translation in linear time. arXiv. Preprint posted online on October 31, 2016 [FREE Full text]
23. Bai S, Kolter JZ, Koltun V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv. Preprint posted online on April 19, 2018 [FREE Full text]
24. Suh Y, Dieterich S, Cho B, Keall PJ. An analysis of thoracic and abdominal tumour motion for stereotactic body radiotherapy patients. Phys Med Biol 2008 Jul 07;53(13):3623-3640. [doi: 10.1088/0031-9155/53/13/016] [Medline: 18560046]
25. Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult. IEEE Trans Neural Netw 1994;5(2):157-166. [doi: 10.1109/72.279181] [Medline: 18267787]
26. Pham T, Tran T, Phung D, Venkatesh S. Predicting healthcare trajectories from medical records: A deep learning approach. J Biomed Inform 2017 May;69:218-229 [FREE Full text] [doi: 10.1016/j.jbi.2017.04.001] [Medline: 28410981]
27. Pascanu R, Mikolov T, Bengio Y. On the difficulty of training recurrent neural networks. 2013 Jun 16 Presented at: The 30th International Conference on International Conference on Machine Learning (ICML); June 16-21, 2013; Atlanta, USA.
28. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997 Nov 15;9(8):1735-1780. [doi: 10.1162/neco.1997.9.8.1735] [Medline: 9377276]
29. He KM, Zhang XY, Ren SQ, Sun J. Deep residual learning for image recognition. 2016 Presented at: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); June 27-30, 2016; Las Vegas, Nevada, USA. [doi: 10.1109/cvpr.2016.90]
30. Kingma DP, Ba JL. Adam: A method for stochastic optimization. arXiv. Preprint posted online on December 22, 2014 [FREE Full text]
31. Goodfellow I, Bengio Y, Courville A. Deep Learning. Cambridge, Massachusetts: MIT Press; Nov 2016.

Abbreviations

LSTM: long short-term memory
RMSE: root mean square error
RNN: recurrent neural network
TCN: temporal convolutional neural network

Edited by R Kukafka; submitted 18.01.21; peer-reviewed by B Li, N Kim; comments to author 17.03.21; revised version received 04.06.21; accepted 05.07.21; published 27.08.21.

Please cite as:
Chang P, Dang J, Dai J, Sun W
Real-Time Respiratory Tumor Motion Prediction Based on a Temporal Convolutional Neural Network: Prediction Model Development Study
J Med Internet Res 2021;23(8):e27235
URL: https://www.jmir.org/2021/8/e27235
doi:10.2196/27235
PMID:34236336

©Panchun Chang, Jun Dang, Jianrong Dai, Wenzheng Sun. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Perspectives on Fruit and Vegetable Consumption and Government Dietary Guidelines: Content Analysis of Comments on News Websites

B Hornsby[1], MSc; H Ensaff[1], PhD

[1] Nutritional Sciences and Epidemiology, University of Leeds, Leeds, United Kingdom

Corresponding Author:
H Ensaff, PhD
Nutritional Sciences and Epidemiology
University of Leeds
School of Food Science and Nutrition
Leeds, LS2 9JT
United Kingdom
Phone: 44 1 133 433 418
Email: h.ensaff@leeds.ac.uk

Abstract

Background: News websites are an essential source of medical news for the public. Many websites offer users the opportunity to leave comments, which may provide rich insights into public perspectives on health issues. With an established role in public health, fruit and vegetable (FV) consumption is central to the government's dietary guidelines. However, FV intake continues to fall short of government recommendations.

Objective: Using comments from news websites, this study aims to explore public perspectives on FV intake and related government dietary guidelines.

Methods: Data comprised 2696 web user comments generated in response to substantial media coverage for a meta-analysis examining FV consumption and the risk of all-cause mortality, cardiovascular disease, and total cancer. Using an inductive thematic approach, the data were analyzed and coded in an iterative process.

Results: Four overarching themes emerged: personal factors, rejection, lack of knowledge, and food landscape, each with component subthemes. The lack of clarity around government dietary health guidelines was apparent, and this, along with emergent personal factors, may hinder better consumption. Rejection was also evident, as was a quality versus quantity of life debate.

Conclusions: There are gaps in the public's understanding of government guidelines, which may act as a constraint to better compliance. Further work should examine this issue and rejection and the possibility of public fatigue related to dietary health information and news. Similarly, future work should also explore targeted interventions with a specific emphasis on health literacy.

(J Med Internet Res 2021;23(8):e19917) doi:10.2196/19917

KEYWORDS

medical news; online news; user comments; public health; population health; qualitative analysis; perspectives; dietary guidelines; diet; fruit and vegetable consumption; mobile phone

Introduction

Background

At present, mass media is popular and influential in society, with 95% of adults reportedly following the news [1]. There is also a growing acknowledgment of the media's role in framing public health issues [2]. Web-based media platforms are an influential news source, and 60% of adults (16-34 years) follow the news through internet or mobile apps [1]. Therefore, this provides an opportunity to explore public perspectives related to public health issues. Specifically, web-based comments in response to media-reported medical research and news can act as a rich source of data for documenting public response. Previous studies have analyzed web-based news comments to explore perceptions related to weight loss surgery [3], dietary risks [4], human papillomavirus vaccination [5], and COVID-19 [6]. Similarly, a previous work [7] has also identified clear advantages with this approach, such as the immediacy of responses that reflect current attitudes (as opposed to recollected opinions). Furthermore, prior studies have reported the value of this approach in assessing public opinion [8] and revealing public perceptions [9].

The consumption of fruit and vegetables (FVs) has been established as a key element in public health dietary guidelines. Indeed, epidemiological evidence highlights FV intake in disease prevention [10-12] and supports its distinction within public health agendas. The UK dietary guidelines on FV consumption focus on *at least five portions of a variety of FVs each day*. This advice is based on the World Health Organization's recommendation of a minimum of 400 g daily to reduce the risk of coronary heart disease, stroke, and some types of cancer [13]. Despite this, the National Diet and Nutrition Survey reveals inadequate intake, with only 31% of adults aged 19-64 years and 26% of adults aged ≥65 years meeting the *5 A Day* guidelines [14]. For children, the levels are worse, with only 8% of those aged 11-18 years meeting the guidelines, with a mean intake of 2.7 (SD 1.5) portions per day [14]. Low FV consumption is also an international issue [15], including among children [16-18].

Multiple models and theoretical frameworks relate to the public's perspectives on FV intake, relevant guidelines, and food choice parameters. In particular, the socioecological model [19] emphasizes the interrelationship between individuals and the environment, as well as the many levels of influence on behavior. The most proximal level comprises an individual's setting and their interactions with those nearest, such as family members, whereas more distal levels can capture interactions and settings, as well as influential social and cultural values and customs. The model captures how a range of aspects can influence an individual's behavior, encompassing factors such as age, gender, income, and the home environment and external influences, such as food policy and retailers. The socioecological model has been used to consider children's FV intake [20], children's obesogenic dietary intake [21], barriers and enablers to healthy eating in young adults [22], FV consumption among low-income groups [23], and food choices of older adults [24]. Another useful theoretical consideration is the theory of planned behavior [25], which has been used previously to understand FV consumption [26-28]. The theoretical framework describes how attitude, subjective norm, and perceived behavioral control drive individual intention, and how intention strongly influences a particular behavior.

Previous research investigating FV determinants has pointed to taste preferences, time, availability, habit, motivation, and knowledge [29-31]. Evidence also points to general awareness for the *5 A Day* message but highlights the need for clearer details on, for example, portion sizes and variety [32]. Similarly, a prior study reported an association between low FV consumption and low knowledge of details pertaining to the *5 A Day* message [33].

Objectives

With the general public's increased consumption of web-based news, user comments can provide a unique data set to explore the public perspectives on FV intake. Indeed, a substantial number of web-based comments were generated following considerable press coverage related to a specific research publication [34]. Aune et al [34] conducted a meta-analysis of 95 studies and found that for all-cause mortality, coronary heart disease, stroke, and cardiovascular disease, the lowest risk was observed with an FV intake of 800 g/day (ie, twice the minimum *5 A Day* UK recommendation). Furthermore, the authors estimated that 7.8 million premature deaths globally in 2013 may be attributable to an intake of <800 g/day, if the associations observed are causal. The aim of this study is to examine web-based comments associated with media coverage (for this meta-analysis [34]) and to explore public perspectives on FV consumption and relevant dietary guidelines.

Methods

Data Collection and Analysis

The top 20 web-based media publications in the United Kingdom (by pageviews) were identified [35]. Foreign language and news aggregator sites were excluded, and the remaining sites were reviewed for web-based articles reporting the findings of previous research [34]. Relevant articles were checked for features that allowed web-based readers to post comments, and these were included in this study.

The news websites had comparable formatting, typified by the news article followed by a comments section, where readers could post their views and opinions. Standard practice dictates that users create an account with a unique username to post a comment. Thus, users can remain anonymous and protect their privacy on the internet. A relevant feature of web-based comments is the ability for commenters to *reply* to others, thereby creating conversations and generating discussions on a specific user-generated topic. Readers without an account, commonly referred to as *guests*, are still able to read others' comments but are unable to post comments. Webpages have standard rules for comments to be published, and users could "report" comments, and those comments deemed offensive or inappropriate were removed by moderators.

All web-based comments for each identified media webpage were sourced, copied (as posted; ie, with spelling, grammar, and expletives unedited), and pasted into a separate document. Some posts were apparently removed by moderators; these were not included in this study. Once compiled, all comments were anonymized, that is, unique identifiers were substituted for usernames and pseudonyms for any identifying details. Anonymized comments were imported into NVivo 11 (QSR International) for data exploration and analysis. Web-based comments were analyzed using an inductive thematic approach. The analysis began with familiarization, where all comments were read and reread. An initial set of themes to capture user comments was created; themes were named appropriately and based on key insights emergent from the early analysis. Next, each source file was analyzed, and each comment was considered sequentially and coded to capture the essence of the comment. The coding was completed for all data to ensure that no comments were lost during the analysis, and all viewpoints were captured. Where data did not relate to any of the initial themes that had been identified during familiarization, then a new theme was considered and added. Thus, all the data were coded appropriately. After this initial round of coding, themes and their respective data were discussed among researchers. Themes were reviewed to ensure that they were a clear reflection of the data. Wherever necessary, themes were modified, for example, some themes merged into others and some secondary themes were created from primary themes. A second round of coding was completed. As a result, some comments were recoded into a newly introduced theme (which more accurately represented the essence of the comment). After this round of coding, themes were once again assessed, discussed, and reviewed before a third round of coding. After this iteration, it was felt that the endpoint of the coding had been reached and the final set of themes represented and described the data well. Researchers undertook several steps to maintain rigor in the analysis; these included scrutiny of the data coding and a detailed discussion of emergent themes. Researchers also practiced reflexivity [36], acknowledging their impact and preconceived perceptions and referencing these during researcher discussions.

Ethical Considerations

Research using qualitative analysis of web-based comments to analyze public perspectives and perceptions is growing. Therefore, the requirements and protocols for this methodology are still being developed. This study was informed by practices in previous studies and ethical considerations for related internet-mediated research [37]. Although commenters would not have been aware that their comments would be used for research purposes, all comments collected in this study were posted on publicly available websites. Furthermore, all comments were anonymized and any identifying words were replaced before the data analysis.

Results

Overview

Overall, 2696 comments (103,930 words) were collected. Table 1 provides the details of the internet-based media publications (page views and readership demographics) and data (number of comments collected and source article headline). It is noticeable that all web-based news headlines referred to eating 10 portions a day. Figure 1 presents a selection of the mastheads and headlines.

Four main themes alongside their component subthemes emerged during data analysis (Textbox 1). Each theme and subtheme are considered alongside representative quotations.

Table 1. Details for the internet-based media publications and data included in this study (N=2696).

Web-based media	Domain pageviews (in millions), n[a]	Readership demographics[b]				Comments, n (%)	Article headline and reference
		Readers (in thousands), n	Male:female	≥35 years:15-34 years	ABC1:C2DE[c]		
BBC News	1715.5	—[d]	—	—	—	175 (6.49)	Fruit and veg: For a longer life eat 10-a-day [38]
The Guardian	331.5	24,363	0.87	1.27	1.80	2369 (87.87)	Forget five a day, eat 10 portions of fruit and veg to cut risk of early death [39]
Mail Online	318.7	26,937	0.79	1.34	1.59	83 (3.08)	Forget five a day, you should eat 10 portions of fruit and veg to cut your risk of early death, researchers find [40]
The Telegraph	169.5	25,464	0.89	1.44	1.66	19 (0.70)	Eat 10 fruit and veg a day for a longer life, not five [41]
Independent	93.9	22,755	0.80	1.41	1.65	17 (0.63)	Five-a-day becomes 10-a-day as scientists urge people to eat more fruit and vegetables [42]
Express	61.3	12,530	1.20	1.64	1.72	1 (0.04)	5? No, you have to eat TEN portions of fruit and vegetables to keep healthy: Latest advice [43]
Yahoo News	56.4	—	—	—	—	20 (0.74)	Now it's TEN fruit and veg a day: Experts issue new guidance for staying healthy [44]
Mirror	55.7	23,910	0.92	1.26	1.52	5 (0.19)	We should be eating TEN portions of fruit and vegetables a day if we want to stay healthy [45]
The Sun	46.2	25,026	0.83	1.06	1.49	5 (0.19)	Doubling fruit and veg to TEN portions per day could save 8 million lives a year [46]
Business Insider	36.7	—	—	—	—	2 (0.07)	Why you should eat 10 portions of fruits and vegetables a day instead of five [47]

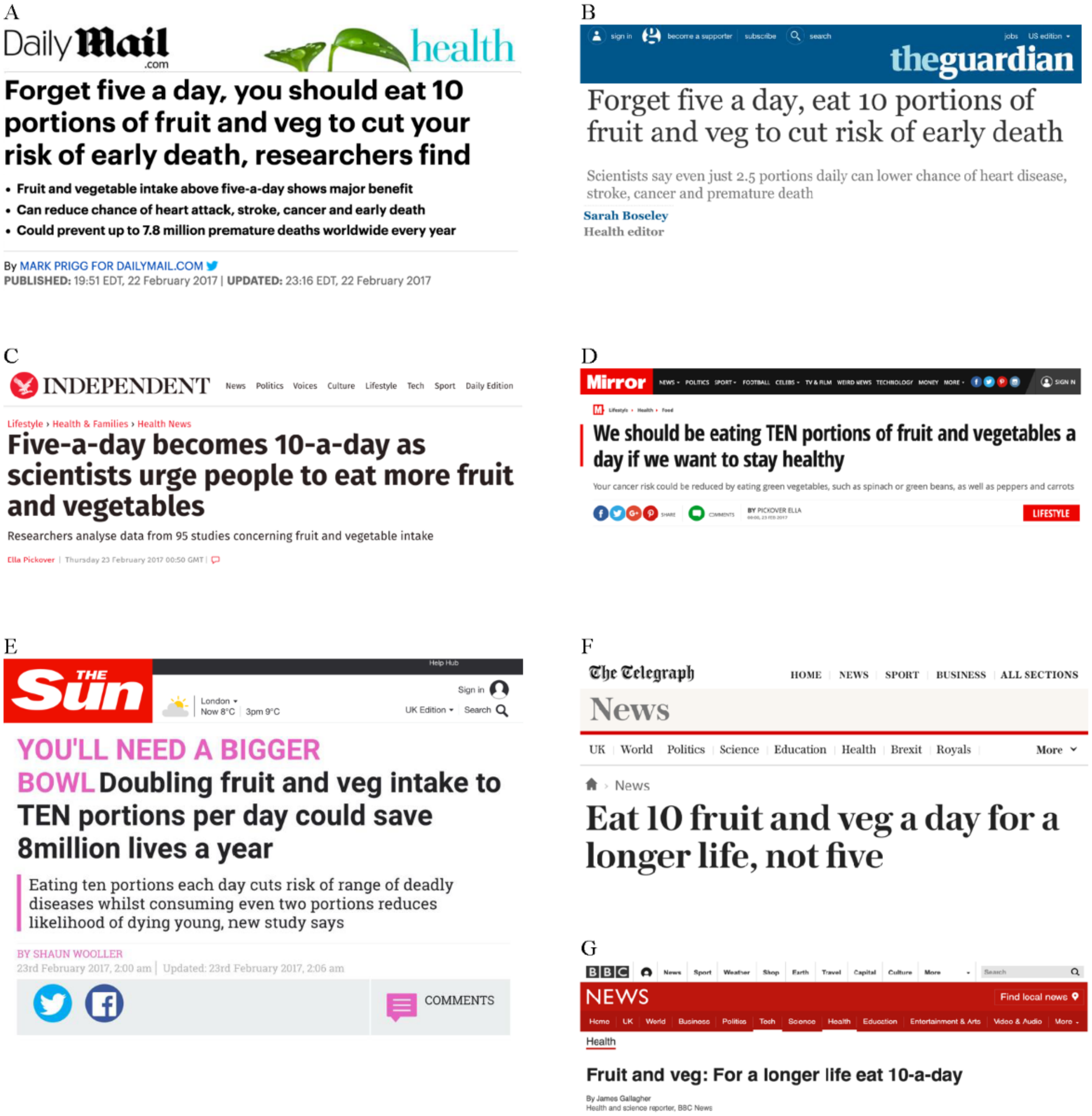
[a] Total pageviews for domains in February 2017 over desktop and mobile [35].

[b] Monthly audience estimates via a PC or via a browser or an app on a device (smartphone or tablet) July 2016-June 2017 [48].

[c] Classification system based on the occupation of the household's chief income earner (ABC1 corresponds to a combination of the more advantaged socio-economic groups, and C2DE corresponds to a combination of the less advantaged socio-economic groups).

[d] Information missing for BBC News, Yahoo News, and Business Insider.

Figure 1. Selection of the mastheads and headlines from the news websites providing comments for this study: (A) Mail Online (image credit: Mail Online); (B) The Guardian (courtesy of Guardian News & Media Ltd); (C) Independent (image credit: ESI Media); (D) Mirror (image credit: Mirrorpix/Reach Licensing); (E) The Sun (image credit: The Sun/News Licensing); (F) The Telegraph (image credit: Telegraph Media Group); (G) BBC News (image credit: BBC News).


A
Daily Mail .com
health
B
sign in
become a supporter
subscribe
search
jobs
US edition
theguardian
C
INDEPENDENT
News
Politics
Voices
Culture
Lifestyle
Tech
Sport
Daily Edition
D
Mirror
NEWS
POLITICS
SPORT
FOOTBALL
CELEBS
TV & FILM
WEIRD NEWS
TECHNOLOGY
MONEY
MORE
SIGN IN
SHARE
COMMENTS
LIFESTYLE
E
THE Sun
Help Hub
London
Now 8°C
3pm 9°C
Sign in
UK Edition
Search
COMMENTS
F
The Telegraph
HOME
NEWS
SPORT
BUSINESS
ALL SECTIONS
News
UK
World
Politics
Science
Education
Health
Brexit
Royals
More
G
BBC
News
Sport
Weather
Shop
Earth
Travel
Capital
Culture
More
Search
NEWS
Find local news
Home
UK
World
Business
Politics
Tech
Science
Health
Education
Entertainment & Arts
Video & Audio
More

A

Forget five a day, you should eat 10 portions of fruit and veg to cut your risk of early death, researchers find

- Fruit and vegetable intake above five-a-day shows major benefit
- Can reduce chance of heart attack, stroke, cancer and early death
- Could prevent up to 7.8 million premature deaths worldwide every year

By MARK PRIGG FOR DAILYMAIL.COM
PUBLISHED: 19:51 EDT, 22 February 2017 | UPDATED: 23:16 EDT, 22 February 2017

B

Forget five a day, eat 10 portions of fruit and veg to cut risk of early death

Scientists say even just 2.5 portions daily can lower chance of heart disease, stroke, cancer and premature death

Sarah Boseley
Health editor

C

Lifestyle › Health & Families › Health News

Five-a-day becomes 10-a-day as scientists urge people to eat more fruit and vegetables

Researchers analyse data from 95 studies concerning fruit and vegetable intake

Ella Pickover | Thursday 23 February 2017 00:50 GMT |

D

Lifestyle › Health › Food

We should be eating TEN portions of fruit and vegetables a day if we want to stay healthy

Your cancer risk could be reduced by eating green vegetables, such as spinach or green beans, as well as peppers and carrots

BY PICKOVER ELLA
00:00, 23 FEB 2017

E

YOU'LL NEED A BIGGER BOWL Doubling fruit and veg intake to TEN portions per day could save 8million lives a year

Eating ten portions each day cuts risk of range of deadly diseases whilst consuming even two portions reduces likelihood of dying young, new study says

BY SHAUN WOOLLER
23rd February 2017, 2:00 am | Updated: 23rd February 2017, 2:06 am

F

News

Eat 10 fruit and veg a day for a longer life, not five

G

Health

Fruit and veg: For a longer life eat 10-a-day

By James Gallagher
Health and science reporter, BBC News

Textbox 1. Emergent themes and subthemes from analysis of comments on news websites.

Personal Factors

- Children
- Taste
- Time, effort, and skill

Rejection

- Ridicule
- Skepticism
- Quality versus quantity of life

Lack of Knowledge

- What counts as fruits and vegetables (FVs), and how much is a portion?
- Limited health knowledge

Food Landscape

- FV cost
- FV availability

Personal Factors

Children

Commenters discussed various influences on FV intake, and within these, several factors emerged: children; taste; and time, effort, and skill. Often, these factors were discussed as barriers, for example, how to overcome difficulties and "to get children to eat much fruit and veg":

> *My kids would tell you they hate mushrooms and don't eat them but they actually eat them all the time without knowing it because I chop them really finely in things like pasta sauce!*

The importance of health education, healthy diets, and FV consumption during childhood was also highlighted:

> *I believe a lot of the problems begins when you're a child. What you pick up from your parents then, determines how you behave later on.*
>
> *If you can't manage it yourself, make sure you feed it to your children from an early age and be a bit tough with them about it. Once they've got through their pernickety hurdles, they'll be healthy veg eaters all their lives.*

Taste

The importance of taste was discussed, with some referring to unpalatable flavors:

> *If vegetables are so good for you then why do they taste so horrible?*
>
> *make fruit and veg more tasty and you got yourself a deal*

In contrast, other commenters spoke positively and recounted food practices:

> *There are few things quite as delicious as a stir fry or a stew with maybe 5 or 6 different veggies in. Trust me, tastes good!*

Time, Effort, and Skill

Commenters also referred to other constraints, including the time and effort required, as well as personal cooking skills:

> *Eating 10 a day won't actually increase your lifespan but it will certainly feel like it with the time taken to source, prepare, eat it...*
>
> *Where the hell do you expect a low-income mum who herself grew up eating frozen lasagna ready meals to summon the wherewithal to buy quite costly fresh ingredients and have time and skill to cook them into tasty meals?*
>
> *I do acknowledge that if you either cannot cook or are very strapped for time, it's much harder to do [eat 10 a day] with shop-bought ready food only. Yet another reason why food education and learning to cook is so important.*

Rejection

Ridicule

A key theme of this study was rejection. This included ridicule, which was mainly referred to through comments relating FV to alcohol and fast food:

> *Does my glass of wine a night count as 1 a day?*
>
> *Grapes are fruit, so drink wine. Hops and barley are vegetables, so drink beer. Cocoa is a plant, so eat chocolate. Cows are herbivores, so eat burgers. Sorted.*
>
> *Too time consuming; a Big Mac followed by a Wagon Wheel and washed down with a can of something is quicker.*

XSL•FO RenderX

Skepticism

Skepticism emerged as a second subtheme within rejection, relating to the reporting, the research itself, and dietary guidelines:

> *No doubt the news tomorrow will be that eating 10 a day gives you cancer.*
>
> *Immediately the media turn this reported research into a requirement to eat 10 portions a day. No wonder everyone is confused or scared by the way health matters are reported.*
>
> *Sorry but this study is nonsense. They asked 65,000 people what fruit and veg they ate in the previous 24h, nothing else about any other time. They then looked at the 4400 who died in the next 7-8 years and predicted the results based on 24 h of their eating. How can that then be used to predict outcomes over a lifetime of eating?*
>
> *First it was eat some fruit & veg every day to live longer, then eat 5x fruit & veg to live longer, now it's eat 10x fruit & veg to live longer, I'm going to just continuously eat fruit and veg (via drip whilst sleeping), I recon I'll be good for the first person to 200!! Unless I get diabetes and need to fast anyway...I wonder what we're going to be told to eat/do tomorrow?*
>
> *5-a-day - plucked out of the air. 10-a-day - plucked out of the air. 28 units of alcohol a week - plucked out of the air. 2 litres of water a day - plucked out of the air. 10,000 steps a day - plucked out of the air. Government health advice - all plucked out of the air.*

Quality Versus Quantity of Life

Interestingly, a debate between the quality and quantity of life emerged:

> *I'm fed up with people harping-on about longevity. Surely the aim of a healthy diet and lifestyle is to improve the overall quality of life, not merely its length.*
>
> *Healthy eating...great I'm all for it. But why the obsession about living longer? I'd rather die at 80 with some dignity, sanity, and independence than live to be 110 as a pale shadow of my former self.*

Lack of Knowledge

What Counts as Fruits and Vegetables (FVs), and How Much is a Portion?

This theme encompassed uncertainty related to FV portion sizes and whether certain foods are counted as FV in the dietary guidelines:

> *They really, really, really need to stop using the word "portions" and "servings". They are vague and mean little to most punters. How many apricots or carrots are there in a portion? How many portions in a bowl of salad? Clearer language is needed in communicating this stuff.*
>
> *I do already eat a lot of mixed nuts. I'm unsure as to whether they count as vegetables, though.*

Limited Health Knowledge

The analysis also revealed a gap in diet and health knowledge more generally:

> *5 or more oranges a day is a huge amount of instantly available fructose and as bad as eating 2 Mars bars daily*
>
> *Mmmm, pesticides and GMOs. Cancerlicious!*
>
> *Ten a day and you will die of pesticides no doubt*

Food Landscape

FV Cost

The wider food landscape emerged as a theme with respect to FV. The cost of FV was highlighted, with comments referring to FV as expensive, citing everyday monetary realities:

> *All well and good but extremely divorced from the reality of everyday life and all its struggles as many cannot afford more then a couple of portions of fruit and veg a day at most*
>
> *...the price of fresh fruit has gone through the roof, the supermarkets are getting beyond a joke now with their pricing. With the govt cutting benefits for everyone who needs them makes veg and fruit unaffordable.*

In contrast, some comments suggest that prudent shopping habits are an effective way to counter the barrier of cost:

> *Go to the frozen section in more or less any supermarket, you can get a 1 kg bag of mixed veg for about 80p.*
>
> *Get down the supermarket and find out the times they yellow label stuff. Just ask, most of them do it at fairly regular times. You can get tons of fruit and veg on the cheap just by lurking at the right times*

FV Availability

The second subtheme within the overarching food landscape is FV availability. This subtheme related to the available quality and variety, for example, *"limited choice, very poor quality"*, and also *highlighted seasonality*:

> *...fruit and vegetables which only kept fresh a couple of days. I Often have to throw away soft fruit with mould after a couple of days.*
>
> *How can we eat 10 portions of seasonal veg without getting bored of turnips and swede?*

Discussion

Principal Findings

The analysis revealed an array of factors relating to public perspectives on FV and reflected the complex nature of food choices and dietary habits. Specifically, FV consumption was affected by the wider food landscape. This resonates with the more distal levels of influence in the socioecological model.

XSL•FO RenderX

Commenters reported that low FV intake was influenced by the produce available in retailers; these comments referred to poor quality, limited variety, and poor shelf life. Similarly, cost was highlighted and there was a general perception that FV is a high-price commodity, with supermarket pricing "beyond a joke." Monetary realities of daily life making FV unaffordable has been reported in previous research in low-income populations, which found that financial considerations impacted FV purchases of parents [49]. Similarly, comments relating to shopping budgets and wages to personal diets reflect the association between income and FV consumption [50], and cost as a barrier to FV intake [27,51,52] as well as associations between food insecurity and FV consumption [53].

A key finding of this study is the personal factors that commenters associated with FV consumption. This included encouraging children to eat FV; notably, parents referred to strategies to increase children's FV intake. Other research has revealed parenting practices and FV availability [54], parents' coping strategies such as playing games with food to make it more appealing [49], and children's growing authority over everyday food [55].

Web-based commenters acknowledged the importance of parents' roles and childhood food experiences. This resonates with individuals' food histories and how previous food encounters (among other factors) contribute to current food choices, as depicted in the food choice process model [56,57]. Parents' recognition of the importance, as shown in this study, is noteworthy. Further research to understand parents' perspectives as food gatekeepers (with chief responsibility for food provision) and, in particular, their strategies and receptiveness to adopting new approaches is warranted. This would add to the valuable research in this area on effective and ineffective parental behaviors [58].

The emergence of time as a potential barrier implies the everyday pressures experienced by many, and the limited and declining time spent on meal preparation [59]. Similarly, this study highlights the effort and skill required. Furthermore, commenters attested to parents' role in imparting cooking skills. Previous research has noted the relevance of food skills in promoting healthier food choices [60] and the relevance of children's food skills [61,62].

One of the key findings of this study is the theme of rejection. This included ridiculing scientific research as well as references to alcohol and fast food. For some commenters, this may reflect an unwillingness to change dietary practices regardless of government health advice; however, further work is required to understand this observation. A noteworthy finding was the debate between quality of life and quantity. Indeed, this implies the potential of healthy eating messages from the perspective of enhancing the quality of life. Further consideration is warranted to establish whether this may be a useful public health strategy to improve dietary intake.

In this study, there was a sense of public messages being unclear and deterring efforts to change dietary habits. An emergent theme was the lack of knowledge, where many commenters expressed confusion related to the current *5 A Day* dietary guidelines. In particular, the portion sizes and what FV qualified were highlighted. This result relays a lack of clarity and a gap in understanding. This also aligns with other work [32,63] reporting confusion related to health messages.

A level of skepticism toward the media was observed from some comments, particularly the reporting of scientific research and health advice. The damage to public trust in nutrition from unrepresentative news stories [64] should not be underestimated. This may contribute to the public fatigue regarding diet and health agendas and underlines the importance of accurate and representative reporting. This is particularly the case, given the pivotal role that mass media play in relaying scientific research.

This study provides valuable insights and contributes to the growing literature using this methodology [3-9,65,66]. The web-based comments provided a rich data source for analyzing public perspectives and perceptions of FV. It is acknowledged that the originators of the data, web-based commenters, are likely to differ from the general population. They have a certain level of web literacy and consume web-based news. Indeed, by definition, they accessed web-based news articles and commented. Interestingly, a cross-national analysis found that 8% of web-based news users in the United Kingdom reported commenting on news websites during an average week [67].

Demographic details of the readership of web-based news media were available. However, the specific details of web-based commenters were not available. Therefore, their demographic characteristics could not be assessed, and this deficit has been acknowledged in similar studies [7,8]. Nevertheless, there is a distinct value in examining these types of data, not least because they do relay the perceptions and perspectives of a portion of the population. Furthermore, some of the value of examining web-based comments can be attributed to the unique features of the data. For example, the anonymity afforded by web-based comments (compared with a research interview) may foster less-filtered data. Indeed, a US study [68] reported that two-third of web-based news consumers felt that anonymity allowed commenters to express ideas that they might otherwise fear expressing. Therefore, although the specific characteristics of web-based commenters are hidden, capturing such perceptions is worthwhile. Web-based comments also have the advantage of being direct responses instigated by users (as opposed to being requested by a researcher). They are within the users' own domain, independent of researchers and without their influence. Similarly, a discussion between commenters may be less constrained, and a relevant feature was commenters *replying* to others' comments, creating a dialog that was beneficial for gathering a range of viewpoints surrounding a topic, for example, FV cost.

Limitations

It is important to acknowledge this study's limitations. Comments were sourced from 10 media outlets. However, most data originated from a single source; therefore, the findings may be specific to this readership. Demographic information for the news media provides some contextual information and may reflect, to some degree, the web-based commenters; however, there is a large degree of uncertainty surrounding the identity and characteristics of the commenters. Some commenters may portray themselves differently on the internet than in *real life*.

For example, previous research has indicated that internet bloggers take on multiple personas across multiple types of social media [69]. Furthermore, given the global reach of web-based news media, all commenters may not have been based in the United Kingdom; indeed, there was evidence of some comments originating from further afield. Automated news commentary [70] is acknowledged, and although all comments in this study were reviewed manually, comment authenticity may have been compromised by bot comments.

Implications for Research and Practice

This study provides valuable insights regarding the public perspectives on FV and related dietary guidelines. Further research combining this approach with more traditional approaches would be beneficial. Similarly, incorporating data from multiple social media platforms is recommended to provide a greater understanding of relevant issues.

This study has implications for health policies and practices, as well as future interventions. Bearing in mind the limitations outlined earlier, findings more broadly point to the need to address the public's health literacy. To some extent, this may also address other barriers, such as time and FV costs. The findings also highlight the need for a better understanding of government guidelines and the need for improved skills. Optimistically, commenters' perceptions of the relevance of food skills were revealed. Incorporating food skills within interventions warrants further investigation. Previous work [71] has identified education through cooking classes can be beneficial for increasing FV consumption. This also resonates with the theory of planned behavior and, specifically, harnesses perceived behavioral control to drive individual intention to influence behavior.

This study highlights the cost and availability of FVs. Given the suboptimal FV consumption (and its place within population health), addressing these factors is vital. Future studies should explore schemes offering discounts to encourage higher consumption. Previous work has indicated that supermarket discount interventions can increase consumer FV purchases and consumption [72]. Furthermore, health interventions may instigate individual-led changes in the home environment. This is commensurate with reciprocal determinism, a concept relevant to the socioecological model, whereby behavior and environment influence each other and individuals can themselves influence the environment.

Similarly, the relevance of parents, as highlighted in this study, needs further consideration. This is particularly relevant for targeted interventions. Finally, the findings related to rejection and skepticism point to the important role that news websites play in informing the public about medical research and news. Furthermore, they underline the importance of representative and accurate reporting of health issues.

Conclusions

A range of factors are relevant in understanding the public's perceptions and perspectives on FV intake. Some relate to the external food environment, whereas others relate to food skills and health knowledge. There is a need to examine the nation's health literacy and its potential role in supporting positive dietary change (specifically FV intake).

Acknowledgments

The authors are grateful to web-based news media (*BBC News*, *The Telegraph*, *The Sun*, *Independent*, *Mirror*, *The Guardian*, and *Mail Online*) for granting permission to use mastheads and images from their websites. The authors would also like to thank Xuehong Li for her contributions to the analysis.

Authors' Contributions

This study was conceived and designed by HE. BH and HE conducted the data analysis. BH wrote the original manuscript. BH and HE reviewed and edited the manuscript.

Conflicts of Interest

None declared.

References

1. News consumption in the UK: 2014 Report. Ofcom : Research Document. 2014. URL: https://www.ofcom.org.uk/__data/assets/pdf_file/0020/52247/News-2014-report.pdf [accessed 2021-05-24]
2. Henderson L, Hilton S. The media and public health: where next for critical analysis? Crit Public Health 2018 Jun 29;28(4):373-376. [doi: 10.1080/09581596.2018.1482663]
3. Glenn NM, Champion CC, Spence JC. Qualitative content analysis of online news media coverage of weight loss surgery and related reader comments. Clin Obes 2012 Oct;2(5-6):125-131. [doi: 10.1111/cob.12000] [Medline: 25586247]
4. Regan A, Shan L, McConnon A, Marcu A, Raats M, Wall P, et al. Strategies for dismissing dietary risks: insights from user-generated comments online. Health Risk Soc 2014 May 28;16(4):308-322. [doi: 10.1080/13698575.2014.919993]
5. Feinberg Y, Pereira JA, Quach S, Kwong JC, Crowcroft NS, Wilson SE, Public Health Agency of Canada/Canadian Institutes of Health Research Influenza Research Network (PCIRN) Program Delivery Evaluation Group. Understanding public perceptions of the HPV vaccination based on online comments to canadian news articles. PLoS One 2015;10(6):e0129587 [FREE Full text] [doi: 10.1371/journal.pone.0129587] [Medline: 26053866]

6. Shorey S, Ang E, Yamina A, Tam C. Perceptions of public on the COVID-19 outbreak in Singapore: a qualitative content analysis. J Public Health (Oxf) 2020 Nov 23;42(4):665-671 [FREE Full text] [doi: 10.1093/pubmed/fdaa105] [Medline: 32648577]
7. Henrich N, Holmes B. What the public was saying about the H1N1 vaccine: perceptions and issues discussed in on-line comments during the 2009 H1N1 pandemic. PLoS One 2011 Apr 18;6(4):e18479 [FREE Full text] [doi: 10.1371/journal.pone.0018479] [Medline: 21533161]
8. Rowe G, Hawkes G, Houghton J. Initial UK public reaction to avian influenza: analysis of opinions posted on the BBC website. Health Risk Soc 2008 Aug;10(4):361-384. [doi: 10.1080/13698570802166456]
9. Lynch M, Brooks-Cleator LA, Giles AR, Rumford MH. "RIP KFC": Public perceptions of a fast-food restaurant closure. Ecol Food Nutr 2019 Jan 27;58(2):120-141. [doi: 10.1080/03670244.2019.1570178] [Medline: 30688089]
10. Zhan J, Liu Y, Cai L, Xu F, Xie T, He Q. Fruit and vegetable consumption and risk of cardiovascular disease: a meta-analysis of prospective cohort studies. Crit Rev Food Sci Nutr 2017 May 24;57(8):1650-1663. [doi: 10.1080/10408398.2015.1008980] [Medline: 26114864]
11. Bazzano LA, He J, Ogden LG, Loria CM, Vupputuri S, Myers L, et al. Fruit and vegetable intake and risk of cardiovascular disease in US adults: the first National Health and Nutrition Examination Survey Epidemiologic Follow-up Study. Am J Clin Nutr 2002 Jul;76(1):93-99 [FREE Full text] [Medline: 12081821]
12. Liu S, Manson JE, Lee IM, Cole SR, Hennekens CH, Willett WC, et al. Fruit and vegetable intake and risk of cardiovascular disease: the Women's Health Study. Am J Clin Nutr 2000 Oct;72(4):922-928. [doi: 10.1093/ajcn/72.4.922] [Medline: 11010932]
13. Diet, nutrition and the prevention of chronic diseases. Report of a Joint WHO/FAO Expert Consultation. Geneva, Switzerland: World Health Organisation; 2003. URL: http://apps.who.int/iris/bitstream/handle/10665/42665/WHO_TRS_916.pdf;jsessionid=7B8112DAB7B1C978530D8864CBD3AA84?sequence=1 [accessed 2021-05-24]
14. NDNS results from years 7 and 8 (combined): data tables. Public Health England and Food Standards Agency. 2018. URL: https://www.gov.uk/government/statistics/ndns-results-from-years-7-and-8-combined [accessed 2020-07-08]
15. Hall JN, Moore S, Harper SB, Lynch JW. Global variability in fruit and vegetable consumption. Am J Prev Med 2009 May;36(5):402-409. [doi: 10.1016/j.amepre.2009.01.029] [Medline: 19362694]
16. Lynch C, Kristjansdottir AG, Te Velde SJ, Lien N, Roos E, Thorsdottir I, et al. Fruit and vegetable consumption in a sample of 11-year-old children in ten European countries--the PRO GREENS cross-sectional survey. Public Health Nutr 2014 Nov;17(11):2436-2444. [doi: 10.1017/S1368980014001347] [Medline: 25023091]
17. Grimm KA, Kim SA, Yaroch AL, Scanlon KS. Fruit and vegetable intake during infancy and early childhood. Pediatrics 2014 Sep;134 Suppl 1:63-69 [FREE Full text] [doi: 10.1542/peds.2014-0646K] [Medline: 25183758]
18. Mihrshahi S, Myton R, Partridge SR, Esdaile E, Hardy LL, Gale J. Sustained low consumption of fruit and vegetables in Australian children: findings from the Australian National Health Surveys. Health Promot J Austr 2019 Jan;30(1):83-87. [doi: 10.1002/hpja.201] [Medline: 30175423]
19. Bronfenbrenner U. Ecological models of human development. In: Postlethwaite TN, Husen T, editors. The International Encyclopedia of Education, 2nd Edition. Oxford, UK: Elsevier; 1994:1643-1647.
20. Ray C, Määttä S, Lehto R, Roos G, Roos E. Influencing factors of children's fruit, vegetable and sugar-enriched food intake in a Finnish preschool setting - Preschool personnel's perceptions. Appetite 2016 Aug 01;103:72-79. [doi: 10.1016/j.appet.2016.03.020] [Medline: 27020091]
21. Mazarello PV, Ong KK, Lakshman R. Factors influencing obesogenic dietary intake in young children (0-6 years): systematic review of qualitative evidence. BMJ Open 2015 Sep 16;5(9):e007396 [FREE Full text] [doi: 10.1136/bmjopen-2014-007396] [Medline: 26377503]
22. Amore L, Buchthal OV, Banna JC. Identifying perceived barriers and enablers of healthy eating in college students in Hawai'i: a qualitative study using focus groups. BMC Nutr 2019;5:16 [FREE Full text] [doi: 10.1186/s40795-019-0280-0] [Medline: 32153929]
23. Robinson T. Applying the socio-ecological model to improving fruit and vegetable intake among low-income African Americans. J Community Health 2008 Dec;33(6):395-406. [doi: 10.1007/s10900-008-9109-5] [Medline: 18594953]
24. Shanks CB, Haack S, Tarabochia D, Bates K, Christenson L. Factors influencing food choices among older adults in the rural western USA. J Community Health 2017 Jun;42(3):511-521. [doi: 10.1007/s10900-016-0283-6] [Medline: 27770374]
25. Ajzen I. From intentions to actions: a theory of planned behavior. In: Kuhl J, Beckmann J, editors. Action Control. Berlin, Heidelberg: Springer; 1985:11-39.
26. Lwin MO, Malik S, Lau J. Association between food availability and young people's fruits and vegetables consumption: understanding the mediation role of the theory of planned behaviour. Public Health Nutr 2020 Aug;23(12):2155-2164. [doi: 10.1017/S1368980019005263] [Medline: 32450936]
27. Jung SE, Shin YH, Kim S, Hermann J, Bice C. Identifying underlying beliefs about fruit and vegetable consumption among low-income older adults: an elicitation study based on the theory of planned behavior. J Nutr Educ Behav 2017 Oct;49(9):717-723. [doi: 10.1016/j.jneb.2017.05.343] [Medline: 28633920]
28. Kothe EJ, Mullan BA. Interaction effects in the theory of planned behaviour: predicting fruit and vegetable consumption in three prospective cohorts. Br J Health Psychol 2015 Sep;20(3):549-562. [doi: 10.1111/bjhp.12115] [Medline: 25209256]

29. Krølner R, Rasmussen M, Brug J, Klepp KI, Wind M, Due P. Determinants of fruit and vegetable consumption among children and adolescents: a review of the literature. Part II: qualitative studies. Int J Behav Nutr Phys Act 2011 Oct 14;8:112 [FREE Full text] [doi: 10.1186/1479-5868-8-112] [Medline: 21999291]
30. Rasmussen M, Krølner R, Klepp K, Lytle L, Brug J, Bere E, et al. Determinants of fruit and vegetable consumption among children and adolescents: a review of the literature. Part I: Quantitative studies. Int J Behav Nutr Phys Act 2006 Aug 11;3:22 [FREE Full text] [doi: 10.1186/1479-5868-3-22] [Medline: 16904006]
31. Guillaumie L, Godin G, Vézina-Im L. Psychosocial determinants of fruit and vegetable intake in adult population: a systematic review. Int J Behav Nutr Phys Act 2010;7:12 [FREE Full text] [doi: 10.1186/1479-5868-7-12] [Medline: 20181070]
32. Rooney C, McKinley MC, Appleton KM, Young IS, McGrath AJ, Draffin CR, et al. How much is '5-a-day'? A qualitative investigation into consumer understanding of fruit and vegetable intake guidelines. J Hum Nutr Diet 2017 Feb;30(1):105-113. [doi: 10.1111/jhn.12393] [Medline: 27334026]
33. Appleton KM, Krumplevska K, Smith E, Rooney C, McKinley MC, Woodside JV. Low fruit and vegetable consumption is associated with low knowledge of the details of the 5-a-day fruit and vegetable message in the UK: findings from two cross-sectional questionnaire studies. J Hum Nutr Diet 2018 Feb;31(1):121-130. [doi: 10.1111/jhn.12487] [Medline: 28573733]
34. Aune D, Giovannucci E, Boffetta P, Fadnes LT, Keum N, Norat T, et al. Fruit and vegetable intake and the risk of cardiovascular disease, total cancer and all-cause mortality-a systematic review and dose-response meta-analysis of prospective studies. Int J Epidemiol 2017 Dec 01;46(3):1029-1056 [FREE Full text] [doi: 10.1093/ije/dyw319] [Medline: 28338764]
35. Sevitt D. UK media publications ranking February 2017. SimilarWeb. 2017. URL: https://www.similarweb.com/corp/blog/uk-media-publications-ranking-february-2017/ [accessed 2020-04-03]
36. Draper A, Swift JA. Qualitative research in nutrition and dietetics: data collection issues. J Hum Nutr Diet 2011 Feb;24(1):3-12. [doi: 10.1111/j.1365-277X.2010.01117.x] [Medline: 21091918]
37. Ethics guidelines for internet-mediated research. British Psychological Society. 2017. URL: https://www.bps.org.uk/sites/www.bps.org.uk/files/Policy/Policy%20-%20Files/Ethics%20Guidelines%20for%20Internet-mediated%20Research%20%282017%29.pdf [accessed 2020-04-04]
38. Gallagher J. Fruit and veg: for a longer life eat 10-a-day. BBC News. 2017 Feb 23. URL: https://www.bbc.co.uk/news/health-39057146 [accessed 2020-04-04]
39. Boseley S. Forget five a day, eat 10 portions of fruit and veg to cut risk of early death. The Guardian. 2017 Feb 23. URL: https://www.theguardian.com/society/2017/feb/23/five-day-10-portions-fruit-veg-cut-early-death [accessed 2020-04-04]
40. Prigg M. Forget five a day, you should eat 10 portions of fruit and veg to cut your risk of early death, researchers find. Mail Online Internet. 2017 Feb 23. URL: https://www.dailymail.co.uk/health/article-4251164/Forget-five-day-eat-10-portions-fruit-veg.html [accessed 2020-04-04]
41. Eat 10 fruit and veg a day for a longer life, not five. The Telegraph. 2017 Feb 23. URL: https://www.telegraph.co.uk/news/2017/02/23/five-a-day-fruit-veg-must-double-10-major-study-finds/ [accessed 2020-04-04]
42. Pickover E. Five-a-day becomes 10-a-day as scientists urge people to eat more fruit and vegetables. Independent. 2017 Feb 23. URL: https://www.independent.co.uk/life-style/health-and-families/health-news/health-news-latest-food-advice-fruit-vegetables-five-a-day-10-portions-a7594631.html [accessed 2020-04-04]
43. No, you have to eat TEN portions of fruit and vegetables to keep healthy: latest advice. Express. 2017 Feb 23. URL: https://www.express.co.uk/life-style/health/770721/Eat-ten-portions-fruit-vegetables-day-keep-healthy [accessed 2020-04-04]
44. Reilly N. Now it's TEN fruit and veg a day: experts issue new guidance for staying healthy. Yahoo News. 2017 Feb 23. URL: https://uk.news.yahoo.com/eating-10-pieces-of-fruit-and-veg-every-day-dramatically-cuts-the-risk-of-cancer-and-heart-disease-101659380.html [accessed 2020-04-04]
45. Pickover E. We should be eating TEN portions of fruit and vegetables a day if we want to stay healthy. Mirror. 2017 Feb 23. URL: https://www.mirror.co.uk/lifestyle/health/should-eating-ten-portions-fruit-9896445 [accessed 2020-04-04]
46. Wooller S. Doubling fruit and veg intake to TEN portions per day could save 8million lives a year. The Sun. 2017 Feb 23. URL: https://www.thesun.co.uk/living/2933247/doubling-fruit-and-veg-intake-to-ten-portions-per-day-could-save-8million-lives-a-year/ [accessed 2020-04-04]
47. Simpson F. Why you should eat 10 portions of fruits and vegetables a day. Business Insider. 2017 Feb 23. URL: https://www.businessinsider.com/why-you-should-eat-10-portions-of-fruits-and-vegetables-a-day-2017-2?r=UK [accessed 2020-04-04]
48. National Readership Survey. 2017. URL: http://www.nrs.co.uk/latest-results/nrs-padd-results/mobile/ [accessed 2020-04-04]
49. Hayter AKM, Draper AK, Ohly HR, Rees GA, Pettinger C, McGlone P, et al. A qualitative study exploring parental accounts of feeding pre-school children in two low-income populations in the UK. Matern Child Nutr 2015 Jul;11(3):371-384. [doi: 10.1111/mcn.12017] [Medline: 23316717]

50. Lallukka T, Pitkäniemi J, Rahkonen O, Roos E, Laaksonen M, Lahelma E. The association of income with fresh fruit and vegetable consumption at different levels of education. Eur J Clin Nutr 2010 Mar;64(3):324-327. [doi: 10.1038/ejcn.2009.155] [Medline: 20087380]
51. Chapman K, Goldsbury D, Watson W, Havill M, Wellard L, Hughes C, et al. Exploring perceptions and beliefs about the cost of fruit and vegetables and whether they are barriers to higher consumption. Appetite 2017 Jun 01;113:310-319. [doi: 10.1016/j.appet.2017.02.043] [Medline: 28263775]
52. Singleton CR, Fouché S, Deshpande R, Odoms-Young A, Chatman C, Spreen C. Barriers to fruit and vegetable consumption among farmers' market incentive programme users in Illinois, USA. Public Health Nutr 2018 May;21(7):1345-1349. [doi: 10.1017/S1368980018000101] [Medline: 29455705]
53. Turnbull O, Homer M, Ensaff H. Food insecurity: its prevalence and relationship to fruit and vegetable consumption. J Hum Nutr Diet 2021 Mar 17:12866. [doi: 10.1111/jhn.12866] [Medline: 33730420]
54. Wind M, Bobelijn K, De Bourdeaudhuij I, Klepp K, Brug J. A qualitative exploration of determinants of fruit and vegetable intake among 10- and 11-year-old schoolchildren in the low countries. Ann Nutr Metab 2005;49(4):228-235. [doi: 10.1159/000087246] [Medline: 16088086]
55. Ensaff H, Bunting E, O'Mahony S. "That's His Choice Not Mine!" Parents' perspectives on providing a packed lunch for their children in primary school. J Nutr Educ Behav 2018 Apr;50(4):357-364. [doi: 10.1016/j.jneb.2017.12.008] [Medline: 29631705]
56. Furst T, Connors M, Bisogni CA, Sobal J, Falk LW. Food choice: a conceptual model of the process. Appetite 1996 Jun;26(3):247-265. [doi: 10.1006/appe.1996.0019] [Medline: 8800481]
57. Sobal J, Bisogni CA. Constructing food choice decisions. Ann Behav Med 2009 Dec;38 Suppl 1:37-46. [doi: 10.1007/s12160-009-9124-5] [Medline: 19787306]
58. Russell CG, Worsley A, Campbell KJ. Strategies used by parents to influence their children's food preferences. Appetite 2015 Jul;90:123-130. [doi: 10.1016/j.appet.2015.02.038] [Medline: 25747856]
59. Appetite for change? Nutrition and the nation's obesity crisis. Kantar Worldpanel. 2013. URL: https://www.kantarworldpanel.com/global/News/Appetite-for-Change-Nutrition-and-the-nations-obesity-crisis [accessed 2020-04-04]
60. Hartmann C, Dohle S, Siegrist M. Importance of cooking skills for balanced food choices. Appetite 2013 Jun;65:125-131. [doi: 10.1016/j.appet.2013.01.016] [Medline: 23402717]
61. Ensaff H, Canavon C, Crawford R, Barker M. A qualitative study of a food intervention in a primary school: pupils as agents of change. Appetite 2015 Dec;95:455-465. [doi: 10.1016/j.appet.2015.08.001] [Medline: 26254277]
62. Ensaff H, Crawford R, Russell JM, Barker ME. Preparing and sharing food: a quantitative analysis of a primary school-based food intervention. J Public Health (Oxf) 2017 Sep 01;39(3):567-573. [doi: 10.1093/pubmed/fdw086] [Medline: 27591297]
63. Hopkins H, van Mil A. BBSRC Public Dialogue: Final Report. 2014. URL: https://bbsrc.ukri.org/documents/1404-food-nutrition-health-final-report-pdf/ [accessed 2020-04-04]
64. Review of nutrition and human health research. Medical Research Council and National Institute for Health Research. 2017. URL: https://mrc.ukri.org/documents/pdf/review-of-nutrition-and-human-health/ [accessed 2020-04-04]
65. Laestadius LI, Caldwell MA. Is the future of meat palatable? Perceptions of in vitro meat as evidenced by online news comments. Public Health Nutr 2015 Sep;18(13):2457-2467. [doi: 10.1017/S1368980015000622] [Medline: 25818555]
66. Thomas-Meyer M, Mytton O, Adams J. Public responses to proposals for a tax on sugar-sweetened beverages: a thematic analysis of online reader comments posted on major UK news websites. PLoS One 2017;12(11):e0186750 [FREE Full text] [doi: 10.1371/journal.pone.0186750] [Medline: 29166399]
67. Kalogeropoulos A, Negredo S, Picone I, Nielsen RK. Who shares and comments on news?: a cross-national comparative analysis of online and social media participation. Soc Media + Soc 2017 Oct 24;3(4):A. [doi: 10.1177/2056305117735754]
68. Jomini SN, Van Duyn E, Peacock C. Survey of commenters and comment readers - center for media engagement. Cent Media Engagem. 2016. URL: https://mediaengagement.org/research/survey-of-commenters-and-comment-readers/ [accessed 2020-04-04]
69. Kurtz LC, Trainer S, Beresford M, Wutich A, Brewis A. Blogs as elusive ethnographic texts. Int J Qual Methods 2017 Apr 28;16(1):A. [doi: 10.1177/1609406917705796]
70. Lokot T, Diakopoulos N. News bots. Digit Journal 2015 Sep 15;4(6):682-699. [doi: 10.1080/21670811.2015.1081822]
71. Brown BJ, Hermann JR. Cooking classes increase fruit and vegetable intake and food safety behaviors in youth and adults. J Nutr Educ Behav 2005;37(2):104-105. [doi: 10.1016/s1499-4046(06)60027-4] [Medline: 15882491]
72. Bernales-Korins M, Ang IY, Khan S, Geliebter A. Psychosocial influences on fruit and vegetable intake following a NYC supermarket discount. Obesity (Silver Spring) 2017 Aug;25(8):1321-1328 [FREE Full text] [doi: 10.1002/oby.21876] [Medline: 28590084]

Abbreviations

FV: fruit and vegetable

Edited by R Kukafka; submitted 06.05.20; peer-reviewed by S Andrews, R Rivera; comments to author 29.06.20; revised version received 18.09.20; accepted 20.01.21; published 19.08.21.

Please cite as:
Hornsby B, Ensaff H
Perspectives on Fruit and Vegetable Consumption and Government Dietary Guidelines: Content Analysis of Comments on News Websites
J Med Internet Res 2021;23(8):e19917
URL: https://www.jmir.org/2021/8/e19917
doi:10.2196/19917
PMID:34420913

©B Hornsby, H Ensaff. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Review

Blockchain Technology Projects to Provide Telemedical Services: Systematic Review

Konstantin Koshechkin[1], PhD; Georgy Lebedev[1,2], PhD; George Radzievsky[1], PhD; Ralf Seepold[1,3], PhD; Natividad Madrid Martinez[1,4], PhD

[1]Federal State Autonomous Educational Institution of Higher Education, I.M. Sechenov First Moscow State Medical University of the Ministry of Health of the Russian Federation (Sechenov University), Moscow, Russian Federation

[2]Federal Research Institute for Health Organization and Informatics, Moscow, Russian Federation

[3]Ubiquitous Computing Lab, HTWG Konstanz, Konstanz, Germany

[4]IoTLab, Reutlingen University, Reutlingen, Germany

Corresponding Author:
Konstantin Koshechkin, PhD
Federal State Autonomous Educational Institution of Higher Education
I.M. Sechenov First Moscow State Medical University of the Ministry of Health of the Russian Federation (Sechenov University)
2-4 Bolshaya Pirogovskaya st.
Moscow, 119991
Russian Federation
Phone: 7 9031516159
Fax: 7 9031516159
Email: koshechkin@gmail.com

Abstract

Background: One of the most promising health care development areas is introducing telemedicine services and creating solutions based on blockchain technology. The study of systems combining both these domains indicates the ongoing expansion of digital technologies in this market segment.

Objective: This paper aims to review the feasibility of blockchain technology for telemedicine.

Methods: The authors identified relevant studies via systematic searches of databases including PubMed, Scopus, Web of Science, IEEE Xplore, and Google Scholar. The suitability of each for inclusion in this review was assessed independently. Owing to the lack of publications, available blockchain-based tokens were discovered via conventional web search engines (Google, Yahoo, and Yandex).

Results: Of the 40 discovered projects, only 18 met the selection criteria. The 5 most prevalent features of the available solutions (N=18) were medical data access (14/18, 78%), medical service processing (14/18, 78%), diagnostic support (10/18, 56%), payment transactions (10/18, 56%), and fundraising for telemedical instrument development (5/18, 28%).

Conclusions: These different features (eg, medical data access, medical service processing, epidemiology reporting, diagnostic support, and treatment support) allow us to discuss the possibilities for integration of blockchain technology into telemedicine and health care on different levels. In this area, a wide range of tasks can be identified that could be accomplished based on digital technologies using blockchains.

(J Med Internet Res 2021;23(8):e17475) doi:10.2196/17475

KEYWORDS

telemedicine; distributed ledger; health information exchange; blockchain

Introduction

In terms of work organization, the health care system has not changed much over the past century. Simultaneously, the technological infrastructure used by people has undergone complete digitalization, with significant changes occurring over the past decades. Blockchain technology has already existed for more than 10 years; however, it did not affect health care in general. Simultaneously, it was clearly shown that telemedicine could significantly affect clinical outcomes in several areas [1].

Interest in using blockchain technology for health care systems has become prominent since 2017. For example, a study conducted in the United States showed high levels of interest in these solutions among medical service consumers. Approximately 19% of the responding hospital executives and 80% of the payers were either considering or were in the process of implementing blockchain solutions [2].

The purpose of this work is to summarize and systematize the methodology for applying blockchain technology in providing telemedicine services. For implementing blockchain technology in telemedicine, surveys were conducted to obtain answers the following questions. How can blockchain-based projects improve telemedicine services? What are the most common features of these solutions?

Telemedical services at the present level of operation allow remote consultations with doctors and consultations of medical workers with patients using communication media such as the Internet. Doctors can monitor patients' health statuses remotely using medical sensors based on the principles of the Internet of Things (IoT). Moreover, doctors can write electronic prescriptions [3,4].

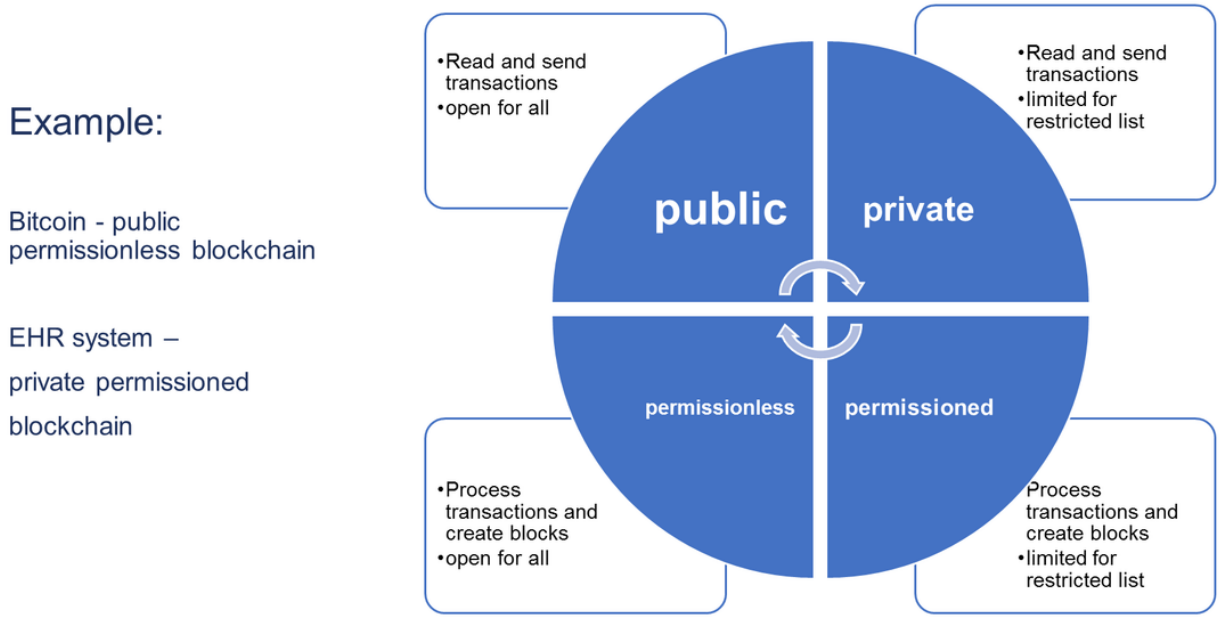
In the future, blockchain technology could potentially help obtain personalized, authentic, and secure health care by merging the entire real-time clinical data of a patient and presenting it in an up-to-date secure health care setup [5,6].

The medical services market is very conservative, and introducing new technologies requires a long time. However, the significant advantages of the data transfer medium and long-term projects can lead to a radical transformation of the health care system as a whole. The provision of telemedicine services can be the main focus of blockchain technology implementation. Solutions for the decentralized distribution of health information constitute one of the most significant benefits of telemedicine [7-9].

Public blockchain is simultaneously a peer-to-peer network and a public database without a central server [10].

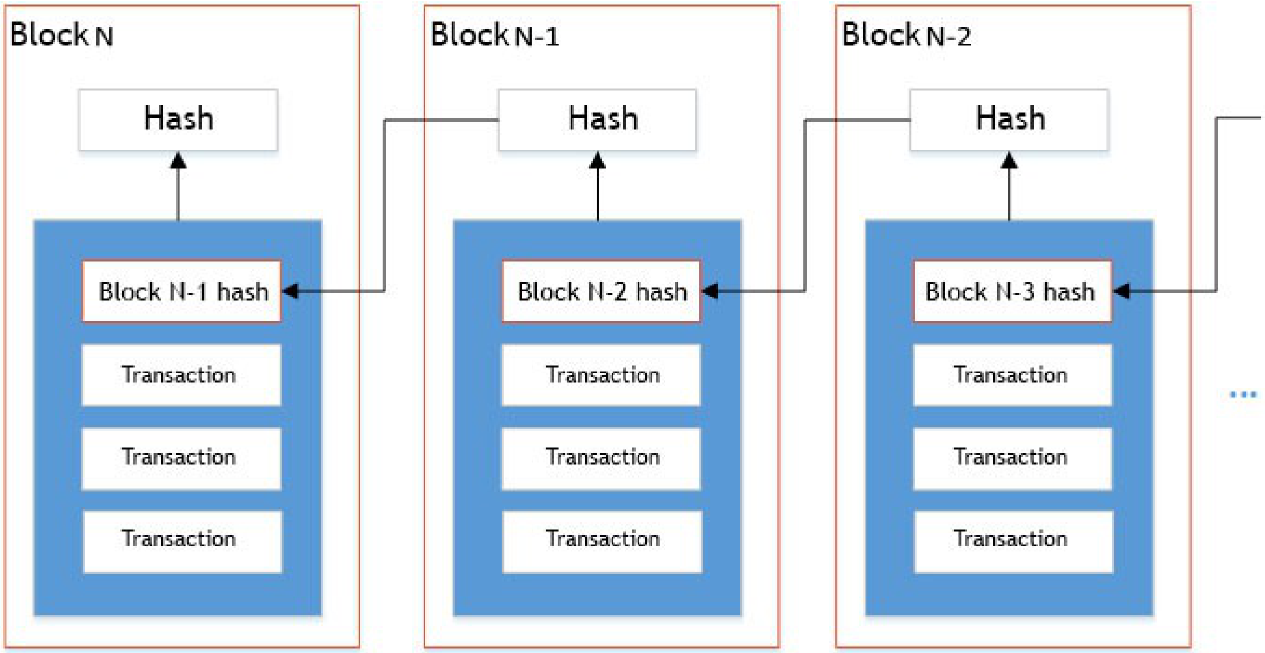
Figure 1 shows the differences between the types of blockchains. In a public blockchain, anyone is free to join and participate in the blockchain network's core activities. A private blockchain allows only the selected entry of verified participants; the operator has the right to override, edit, or delete the necessary entries on the blockchain. A permitted blockchain has the properties of private and public blockchains. Permissioned blockchains have seen an increase in popularity thanks to their ability to allocate specific permissions to various network users. A blockchain database contains information about each transaction of exchange between users. The transactions are verified by the miners, who check the authenticity of the committed actions and then form blocks from these transaction records. Information is distributed on each network member host or the so-called node.

Figure 1. Blockchain types. EHR: electronic health record.


Example:
Bitcoin - public permissionless blockchain
EHR system – private permissioned blockchain
•Read and send transactions
•open for all
•Read and send transactions
•limited for restricted list
public
private
permissionless
permissioned
•Process transactions and create blocks
•open for all
Process transactions and create blocks
•limited for restricted list

As Figure 2 shows, a blockchain is a distributed database with a sequence of attached and attaching blocks where every following block includes the value of the previous block's hash function as hash information. All peers of the peer-to-peer network performing information exchange processes have the same sequence of blocks. A long chain called a transaction log is the result of block interconnections [11].

Figure 2. Blockchain hash function.



Transactions are connected into blocks, where each block contains nearly multiple transactions that form a hash tree. Further, the block is transmitted to a distributed ledger network where it is checked by specified system participants to avoid mistakes and guarantee correctness.

Providing increased attention to distributed registry technology will lead to an understanding of this technology's potential application in health care systems. Such conditions facilitate the integration of blockchain technology into existing projects, the development of new high-tech ones, and working with a large amount of data [12,13].

Cryptographic tokens are programmable digital units of the value recorded on a distributed ledger protocol such as a blockchain. Tokens do not have their own blockchain but depend or exist on an existing blockchain. Tokens may represent fungible or in-fungible units of value in the form of money, coins, points, digital items, or representations of real-world physical items and rights. Tokens can exist on public/open and permissionless blockchains that anyone on the Internet can view, or they can remain private, such as within an enterprise business network.

The last decade has had a revolutionary impact on the way information is handled. More recently, meaningful information was considerably expensive, and the means for its analysis and dissemination were not available to ordinary citizens. Today, information resource availability has significantly increased, which could be a valuable resource for teamwork and decision-making. Blockchain could significantly benefit telemedicine by managing medical data access and teamwork in organizations. Despite their power, tools based on blockchain are easy to use and do not require high financial costs. The only remaining problems are data interoperability and access restrictions due to imperfect regulatory policies, lack of technical regulation in the field of medical data exchange, and deeply rooted paper-based work traditions in many countries [14-17].

At present, it is possible to represent any information, including numerical data, texts, audios, and images, in a digital format suitable for storage and processing on a personal computer; moreover, the necessary infrastructure for rapid distribution of information is available. For example, we use high-speed mobile Internet technology based on long-term evolution, a wireless broadband communication standard. These technologies make it possible to provide instant access to a vast range of information systems using hand-held personal computers, the functionality of which is contained in most modern mobile phones [18].

There is also a widespread introduction of technologies that will make using digital information ubiquitous. However, in health care, it is significantly inferior compared to other areas, such as banking solutions or retail selling of consumer products. Thus, many promising opportunities for the application of information and communication technologies allowing for increased availability and quality of medical services are currently not implemented to the appropriate extent.

Several reviews on digital technologies to change the health care system landscape have been published lately. These reviews state that blockchain technology offers a platform that could be used for many potential applications in health care. Although this technology is in the early stages of design and development, many organizations have proposed solutions that have the potential to increase health care data transparency and operating efficiency. However, the scalability, security, and cost-effectiveness of blockchain technology will require further research prior to large-scale production and deployment [19-21]. The deployment of blockchain technology in telehealth and telemedicine technology is still in its infancy. Several challenges and research problems must be resolved to enable the widespread adoption of blockchain technology in telehealth and telemedicine systems. The latest survey on telehealth and telemedicine systems shows that they are centralized and fall short of providing the necessary information security and privacy and there is a lack of blockchain-based health care studies considering its use in telemedicine [22]. This topic has attracted considerable interest in the context of the COVID-19 lockdowns. The vulnerable situation created owing to the COVID-19 pandemic has shown the necessity of developing a single-source blockchain-based pandemic health record management system to address several existing and future challenges. Storing, sharing, and accessing COVID-19 pandemic data in a single source of information (database) through blockchain is the most crucial step to address the previously stated challenges; nevertheless, many issues need to be resolved by international health organizations, country leaders/governments, and international policy makers to introduce government-to-government digital health service–related policies, data sharing acts, and health policies. Further, issues regarding digital connectivity, digital inequality, and the digital divide that exists primarily in the least- and under-developed countries around the world must be addressed. This pandemic situation is the perfect opportunity for humanity to bring all countries together regardless of their differences under a single umbrella for ensuring world health safety and fighting against COVID-19 and future pandemics [23-25]. In many cases, remote consultations have become the only option for patients. Our study concerns the practical aspects of implementing available blockchain technology in telemedicine.

Methods

Knowledge Review

Our goal was to review the narrow segment of scientific and public domain knowledge regarding the interconnection between telemedicine and blockchain technology. Many successful projects do not have any grounding in research (for example, HapiChain, DocCoin, and others); hence, we decided to include conventional web search engines, where people generally showcase their innovations, often based on personal needs. Most companies in the blockchain space are start-ups based on emerging technologies, whereas the literature is stale owing to the publication lag time (eg, 6-12 months to conduct the research followed by another 3-18 months to publish the findings). In this study, we discovered different types of solutions with similar standard features.

Search Strategy

The search was based on 2 main source types. The first source was web-based journal databases, indexes, and reference lists. We searched for prototypes and worked in progress using the following search terms: telemedical, telemedicine, blockchain, and distributed ledger. We constructed a search string using AND, the conjunction logical operator, and OR, the disjunction logical operator ([telemedical OR telemedicine] AND [blockchain OR distributed ledger]). The search was based on the metadata, including, title, abstract, and keywords. We targeted original research papers and review articles indexed by PubMed, Scopus, Web of Science, IEEE Xplore, and Google Scholar.

The second source comprised conventional web search engines (Google, Yahoo, and Yandex). The first 100 results from each search engine were analyzed as the most relevant results to find the websites of the relevant projects. We performed a search for whitepapers, system manuals, relevant user information, and other information describing project features on these websites. We searched the web-based journal databases and websites independently of each other. We searched the journal databases first and then searched the related sites.

Selection Criteria

The main inclusion criterion was that the project must play a role in telemedical service and use blockchain technology as a significant domain. We settled on this criterion to filter out projects based on the opinions of their developers. From a broader perspective, authors can suggest many blockchain-based applications in theory that are suitable for telemedicine services. Our approach was to use only the data about those projects being used in telemedicine. We excluded projects without English-language websites. We also excluded projects without a definite whitepaper or similar information guide. Projects without precise contact data consisting of addresses and phone numbers were also excluded. Publications for the last 10 years only were considered. At least 2 authors checked all the selected projects. Data on similar projects found from different sources were combined.

Results

Evaluation and Assessment of Project Features

We analyzed the following features: medical data access (electronic health record [HER] distribution), medical service processing, video conferencing, epidemiology reporting, diagnostic support (with artificial intelligence [AI] technology), treatment support (with AI technology), patient data aggregation (for clinical trials, etc), visit arrangements for medical procedures, ordering medicines from pharmacies, payment processing, and fundraising. These features resulted from the analysis performed by the coauthors on assessing the cross-represented systems. Their significant features were identified, and similar features in different systems were found.

We assessed the suitable projects and recorded their functionalities in a spreadsheet (see Multimedia Appendix 1). We analyzed the function descriptions in the published articles and project website materials.

Data Analysis Results

We found 37 matches in PubMed, 8 publications in Scopus, 5 in Web of Science, 22 publications in IEEE Xplore, and 547 results in Google Scholar. As for the web search engines, we obtained 118,000 results from Google, 242,000 from Yahoo, and 118,000 results from Yandex. From the first 100 web pages of each search engine and the most relevant results from journals, a total of 36 suitable projects were identified (only 3 of them in journal databases). As the results show, all the identified projects can be divided into the following areas: tracking the origin of data (2/36, 5.6%), storing and managing data (21/36, 58%), telemedicine services (5/36,14%), diagnosing (3/36, 8%), and using blockchain to raise funds (5/36, 14%).

We can describe these areas as follows: The origin of data is an implementation of the authentication procedures based on blockchain. Storing and managing data could be possible using blockchain solutions in different ways. Generally, the most sensitive data could be stored in blocks on the blockchain itself. However, this is almost impossible for significant amounts of data owing to computational difficulties. The best solution is to store data on the cloud, in an encoded manner, with the data key stored in the blockchain. Software solutions could be limited to only the authentication and data transfer provision or could perform as telemedicine service providers themselves with voice- and video-streaming capacities. Another way in which blockchain could help to improve the health care system is through smart contracts. They could be used for automation and control of the diagnostics. If we have a list of procedures to be performed before we are sure of the examination results, we could load this list into a smart contract and perform them.

Moreover, we cannot ignore the financial aspect of this technology. As the popularity of blockchain stems from Bitcoin and other cryptocurrencies, it is used in many cases to sell tokens as a representative of some value. These tokens are proposed to be used to pay for medical services, and they could be earned for clinical trial participation or adherence to the prescribed treatment.

Selected Projects

Only 18 projects met the selection criteria. The 5 most prevalent features of the available solutions (N=18) were the following: medical data access (14/18, 78%), medical service processing (14/18, 78%), diagnostic support (10/18, 56%), payment transactions (10/18, 56%), and fundraising for telemedical instrument development (5/18, 28%).

The primary role of digital systems based on blockchain technology for telemedicine is to distribute the medical information of patients and provide access to this data for specialists. Usually, the systems working in this direction allow the patient to control which medical data are to be placed in the system and who can access these data. As the implementation of access control does not require any unique technical means or bureaucratic delays, the possibilities that open up to users are extensive [26].

In this study, we examined several examples of currently existing telemedicine systems based on blockchain technology. The Medicalchain project (United Kingdom) mediates patients' consultations with doctors. Medicalchain is based on a double blockchain structure: The first blockchain controls access to medical data and is built using Hyperledger Fabric, whereas all applications and services on the platform run on Ethereum. Any interactions with medical records are recorded as transactions in the blockchain registry [27].

The Symptomatic Platform (United States) was created to support patients with multiple sclerosis, but it is suitable for managing data on any chronic disease. Telemedicine in this system is carried out by video conferencing. Users can store, analyze, and compare their data with others; they can also submit reports, perform statistical evaluation of the epidemiological situation, and undergo genetic screening [28].

The Docademic system (Mexico) facilitates patients' communication with doctors through videoconferencing. Medical data are stored in the blockchain system. For doctors, the system offers reports on epidemiology, tips on diagnosis, and treatment suggestions. In addition, experts have the opportunity to interact with large groups of patients with similar disease profiles and other characteristics. Furthermore, it can be used by patients to pay for services using cryptocurrency [29].

Robomed (Russia) is controlled and administered by smart contracts based on the Ethereum blockchain. The system aims to provide telemedicine services to patients. The Robomed electronic health record (her) medical information system allows medical organizations to register, connect, and operate on the Robomed network. Robomed's core functions include real-time monitoring of all patient interactions, health care staff decision-making, access privileges, health care professional scheduling, patient health analysis, and consulting services. The Robomed mobile app allows patients to receive telemedicine consultations and exchange EHRs. Using smart contracts, Robomed organizations can track and verify patient health statuses and adhere to clinical guidelines for health care services [30-32].

The DocCoin project (Estonia) has a similar working principle. The user receives access to the service through the “Doc in pocket” mobile app. In the system, doctors receive payment for their services in electronic currency. The system's development began in 2015, with an initial coin offering in 2018 to exchange its tokens for cryptocurrency. Payment is charged for services such as storing medical data, visiting specialists, ordering medicines, and visiting medical institutions [33]. DocCoin is a global service that integrates the entire online medicine industry and offers advantages to businesses and clients. DocCoin provides access to doctors around the world through its smart contracts. Every user can receive specialist advice 24/7 anywhere in the world in any language.

The Memorial Hermann Health Network network helps connect patients and health care organizations, providing them with the ability to independently manage and control their data. This system enables individual organizations to provide data anonymously for scientific research. The system is implemented on the basis of blockchain technology and allows one to create an account for storing personal data; smart contracts provide encryption, and this system is aimed at big data analytics. A blockchain-based solution helps distribute control among stakeholders, which provides strong protection against fraudulent activities. Blockchain also provides transparency, traceability, auditing, and security, and allows data to be identified through decentralized storage of information [34].

The Trusted Health system (United States) has a similar operating principle, currently preparing for the release of its tokens. The accumulated medical information can be used for scientific research. The most significant interest in patient data is expressed by organizations conducting clinical studies, insurance companies, pharmaceutical companies, and organizations providing analytical and consulting services. The presented information about the health of patients helps specialists improve treatment methods, attract clients, increase profits, and reduce expenses while simultaneously reducing the cost of collecting information, thus improving its quality [35].

Another application of blockchain technology is the creation of diagnostic systems. For example, the Skychain project (Russia) is an infrastructure for the placement, training, and use of AI, designed to perform diagnostics. As planned, AI should not replace the doctor but must monitor medical decisions to prevent mistakes. It is supposed to use its cryptocurrency to function, which is necessary to run smart contracts. The proceeds will be received by the owners of the neural networks performing diagnostics. Doctors and patients will be able to evaluate the results of research by several neural networks simultaneously and verify the diagnosis based on the obtained data [36]. This can be done without personal visits to different specialists. Once created, medical images or other diagnostic materials can be evaluated by distant doctors using telemedical services.

The DeepRadiology (United States) project aims to train machine intelligence but in a narrower direction. Medical images obtained by radiological methods are taken into account. This decision is relevant owing to the high degree of subjectivity in evaluating the medical images. The result of the study depends entirely on the qualifications and experience of the specialist evaluating the image. In November 2017, DeepRadiology reported on the first artificial intelligence (AI) system that could interpret computed tomography (CT) with a performance level higher than that of doctors. The system was trained using more than 9 million brain tomography images. The blockchain technology in this system acts to implement a secure storage medium [37]. Blockchain would most effectively integrate as a mode of managing access to sensitive health data. By storing an index of health records and the related metadata linked to the sensitive data (stored elsewhere on a secure cloud), the system would introduce a layer of interoperability to the currently disjointed set of systems.

The eHealth First platform is an international project with the goal of implementing natural language processing as most nonstructured medical records filled out manually. For these purposes, solutions based on AI are also applied. The rest of the system is similar to the previously described ones. It is based on the medical data of patients stored using blockchain technology. It allows associating patients with specialists for diagnosis and consultation. The system also offers researchers solutions, allowing them to conduct research based on accumulated data [38].

In another international system, CareX (United States, Canada, and India), the main emphasis is placed on using the solutions of this platform as a means of international payments. In some cases, the financial transfers through bank payments, even for the payment of medical services from one country to another, raises questions from the financial supervisory authorities in the sending and receiving countries. The remaining functions of the transmission of medical information in this system are also presented. An innovative aspect of this project is the ability to communicate with the chatbot, which implements AI elements for making a preliminary diagnosis [39].

Similar issues affect the operation of the Solve.Care system (Estonia and Ukraine). It allows medical organizations, insurance companies, and patients to work without intermediaries to guarantee payment and provision of services. They are using their tokens for payment [40]. In this group of systems, blockchain technology serves as a supporting service for telemedical interactions.

HealPoint (United States) is based on the Ethereum platform for the implementation of telemedicine services. The system helps patients use medical consultation services, transfer symptoms, medical records, and vital signs of the patient. Ethereum-based smart contracts allow patients to obtain a second opinion from several medical experts around the world. Before providing medical services, network specialists verify the doctor's identity and license. All interactions with patients are digitally signed before being recorded on the blockchain for audit purposes [41].

HapiChain (France) exploits blockchain technology to improve the security, scalability, and reliability of medical workflows. Although HapiChain is patient-centric, it also helps clinicians save time and prevent unnecessary trips without improvising treatment. In HapiChain, two primary telemedicine services are embedded, namely telemonitoring and teleconsultation. HapiCare, an existing health care monitoring system with self-adaptive coaching using probabilistic reasoning, is used for telemonitoring. HapiChain then completes this service by adding teleconsultation services exploiting blockchain technology [12].

High interest in solutions based on cryptocurrency implemented in blockchain technology causes widespread projects to attract investment. This approach allows researchers to fully maintain control without transferring the right to make decisions to a critical investor, board of directors, or any other collegial body formed by financing participants. For example, Elcoin (Russia) aims to obtain funding to develop medical equipment through the release of its tokens. The company has created several medical devices. Funds are needed for further development, which the company collects by electronic emission of tokens. The primary target audience is foreign investors [42]. Therefore, blockchain technology can be used as a medium for fundraising to develop telemedical services.

Technical solutions are also used for the noninvasive diagnosis of diabetes, stomach ulcers, and lung cancer. The Health Monitor project (Czech Republic) is based on the principle of analyzing markers of biochemical processes in the body; in a solution created by system developers, the marker is a mixture of gases exhaled by humans [43]. Furthermore, this research project primarily uses cryptocurrency-based infrastructure as a tool to search for investments.

More unusual is the PointNurse system (United States), which is primarily aimed at telemedicine services, establishing a visual link between specialists and patients. However, it focuses mainly on nursing staff, which is its singular goal. It allows practicing nurses and members of the support team to conduct direct consultations on primary health care to make a health assessment. The system also has its cryptocurrency used to pay for services and encourage patients to follow the recommendations. The system works with specialists in several languages. There is a rating for specialists, and patients can choose whom to contact [33,44,45].

MedCredits (United States) uses Ethereum to help doctors diagnose dermatological patients using telemedicine. It is a secure system that protects users from intruders through the implementation of reputation-based systems. Moreover, it allows one to check doctors' licenses. Two Ethereum-based smart contracts implemented in MedCredits help automate escrow-protocol-based payments and validate medical services. The doctor can access the patient's symptoms to diagnose and prescribe treatment using the blockchain [33,46].

The common feature of blockchain-based services is that they deal with sensitive private data that can affect patients' health. Therefore, security and privacy issues should be solved first. Blockchain technology–based systems can quickly provide suitable solutions, as observed in clinical trials [47].

Discussion

Telemedicine Services

Telemedicine services, including radiology, dermatology, and cardiology services, could be provided in different ways to help chronic patients and monitor acute patients. Real-time telemedicine (also called live telemedicine) makes it easy to facilitate doctor-patient interactions anytime and anywhere. Live telemedicine includes videoconferencing and telephonic consultations that let providers and patients communicate in real time. Assessments of medical histories, essential visual examinations, psychiatric evaluations, and even ophthalmic tests can all be conducted via real-time telemedicine. Remote patient monitoring allows health care providers to monitor patients' health data remotely, usually when they are in their own homes. Remote patient monitoring is especially effective for chronic conditions, ranging from heart disease to diabetes to asthma. Telemedicine can improve communication among the members of a medical team. A primary physician can get greater access to a wide range of specialists without requiring any travel. Telemedicine technology has considerably accelerated the rate at which X-rays, computed tomography scans, and other important images are distributed from one medical professional to another. Thanks to telemedicine, health care professionals have multiple ways to interact with patients in their own homes. Web-based services, such as patient portals, allow providers to share essential information and answer simple questions. Telemedicine could be the only option in harsh conditions; for example, space stations and polar expeditions could receive only remote professional medical help. The outbreak of COVID-19 proved that telemedicine was an effective option to fight a pandemic [24].

For example, a patient can obtain advice from a specialist without visiting another clinic or making a trip to a remote city or country. There is no need to remain on the waiting list; systems with a large pool of doctors instantly allow patients to find the required specialist. The expansion of such systems to the financial sector can also be implemented through blockchain transfers using cryptocurrencies. At the same time, within the framework of the smart contracts, participants are guaranteed the provision of services on the one hand and payment on the other. Smart contracts are pieces of code that sit on the blockchain. Once a smart contract is executed, payments are automatically deducted from a patient's digital wallet, and funds are moved into the supplier's digital wallet; therefore, payment happens seamlessly. Inside a smart contract, exact rules could be included, describing the conditions such as when payments will be transferred. It could be an acknowledgment of some medical data or procedure completion.

Smart Contracts and Tokens

Blockchain technology can facilitate such an infrastructure in the form of a decentralized marketplace where access to health data is under the individual's control. Information seekers can post their queries and individuals can remain anonymous and decide whether they want to share their data. With the tokens in a blockchain-based marketplace, a reward can be automatically transferred based on a digital contract once the data has been delivered. Such a system has a clear advantage over a fiat currency–based system where an agent must always be involved, and the large population of unbanked individuals cannot participate.

Diagnostic centers can be included in the same unified system; following the recommendation of a physician made via a telemedical service system, an appointment can be scheduled for a specialized examination using inpatient diagnostic equipment that requires an on-site visit. The same principle applies to working with pharmacies. Pharmacy institutions connected to the system will receive a guarantee that this drug was prescribed by the physician who referred to the patient indicated in the system. The high degree of transaction security in blockchain systems ensures a suitable level of data authenticity.

Smart contracts are extremely useful for telemedicine. They are specific computer codes built into the blockchain network and are executed on computers or nodes. Terms between the parties in the smart contract are written in the form of code in the blockchain. The involved parties are anonymous, but the contract itself has available public properties. When a starting event has happened, for example, the occurrence of a specific date, the contract launches itself based on the conditions of the provision recorded in its code. Following the terms of the contract, the network nodes update the register. After all the requirements are met, the contract is automatically closed, and information about the actions performed is recorded in the blockchain.

In this work, 18 blockchain systems for telemedicine were analyzed. The results are presented in Multimedia Appendix 1.

Their capacities were summarized and systematized under a common methodology to identify suitable opportunities for applying blockchain technology to provide possible telemedicine services. To determine the required properties, the role of telemedicine services in the health care system has been evaluated. Furthermore, the role of digital systems based on blockchain technology for telemedicine has been determined. The most promising projects available on digital support of telemedicine services using blockchain technology are presented in this paper.

Analysis of existing solutions has shown advantages for patients, medical organizations, and related institutions. Telemedicine solutions can increase the availability of medical services for patients, reduce the burden on medical institutions, reduce the cost of providing services, and increase their delivery efficiency. Telemedicine does not require expansive facilities and can optimize the utilization of on-site medical staff resources. Blockchain technology addresses the issue of access to medical data and the preservation of their confidentiality. Another reason for blockchain system implementation is its financial capacity. It will ease the transfer of payments for medical services from one country to another without bureaucratic delays. It could be used as a payment method from one medical institution to another, from a hospital to a foreign bank, or from a patient to a foreign hospital. Nevertheless, all participants need to have blockchain system tools to achieve this objective or employ brokers to convert blockchain tokens to fiat money. In addition, governmental control and audit of foreign exchange inflows and outflows will be needed.

Blockchain and Telemedicine Drawbacks

Despite general concerns against blockchain as with any new technology, these solutions are successful at the pilot level. If we try to address only the most common issues, namely decentralization risks, expenses, and computational complexity, blockchain can become a unified (but decentralized) medical information database. In future, it will be possible for doctors to fill this database not only from EHRs but also with Internet of Things gadgets. The blockchain can also record the results of group examinations from diagnostic centers and information about clinical trials of drugs. Any hospital in the country will have authorized access to patient data. This will enable doctors to share clinical trial results faster, which will accelerate the development of drugs for severe diseases. Centralization in this perspective will create eminent dependency on a central node. This will result in a bottleneck for the entire system.

In telemedicine, the main issue is medical data protection. Additional effort is required to prevent unauthorized access to patients' medical data, especially if they access telemedicine on a public network or via an unencrypted channel. One of the main challenges faced by blockchain solution developers is securing the blockchain. For example, in Bitcoin, miners use their computing systems that consume electricity to perform the computations required to verify data on the blockchain. In return, they receive a reward in the form of digital money.

Projects not related to cryptocurrencies cannot provide network participants with a similar form of reward. Therefore, it is difficult for them to generate public interest. In telemedicine systems, the information in the blockchain can be protected by medical organizations and research agencies. They will act as miners and use their hardware and computing infrastructure to maintain the integrity of the blockchain. In return, the system will provide them with access to anonymous patient data (with the latter's consent) for epidemiological and other research activities.

For example, such a project has already been implemented in Estonia [48]. Since 2008, all hospitals and doctors in the country have been mandatorily digitizing health information. Recently, blockchain has become responsible for its own security.

The distributed ledger acts as a database for medical records. When changes are made to a patient's medical record, this event is immediately recorded on the network, along with information about what exactly was changed, deleted, or added. This transparency level allows therapists, surgeons, pharmacists, and other professionals to receive up-to-date and correct patient information acceptable to the community.

The system makes it possible to make more accurate diagnoses, taking into account a fully documented medical history, guide doctors in crises (for example, provide information about allergic reactions to drugs), and even adjust treatment for chronic diseases over time (depending on changes in the patient's condition).

Conclusions

Our study identified several blockchain technology projects to provide telemedical services. They differ considerably from each other and should be examined separately. There is no “silver bullet” in this market, but we suggest some points of interest to start with in our review. These various features (medical data access, medical service processing, epidemiology reporting, diagnostic support, and treatment support) allow us to talk about the possibilities of integrating blockchain technology into telemedicine and health care on different levels. In this area, a wide range of tasks can be identified that could be accomplished based on digital technologies using blockchains. Almost all existing projects are prototypes that are under development. Considering that many of them are aimed at attracting investments, it can be said that the role of scientific, technical, legal, and economic experts in assessing the feasibility of investing in various start-ups based on blockchain technologies is increasing.

Solutions based on blockchain technology are of most interest for countries where there are currently no working centralized information processing systems used for telemedicine. Launching a distributed system based on open-source solutions provides a potential opportunity to avoid significant investments in building a centralized system.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Functionalities of the suitable blockchain-based projects analyzed in this study.
[XLSX File (Microsoft Excel File), 12 KB - jmir_v23i8e17475_app1.xlsx]

References

1. Hersh WR, Helfand M, Wallace J, Kraemer D, Patterson P, Shapiro S, et al. Clinical outcomes resulting from telemedicine interventions: a systematic review. BMC Med Inform Decis Mak 2001 Nov;1(1):5 [FREE Full text] [doi: 10.1186/1472-6947-1-5] [Medline: 11737882]
2. Healthcare industry interest in blockchain is heating up. Q3 Black Book Blockchain Vendors. 2017. URL: http://blackbookmarketresearch.com/uploads/pdf/2017 [accessed 2021-04-01]
3. Martinez AW, Phillips ST, Carrilho E, Thomas SW, Sindi H, Whitesides GM. Simple telemedicine for developing regions: camera phones and paper-based microfluidic devices for real-time, off-site diagnosis. Anal Chem 2008 Apr;80(10):3699-3707. [doi: 10.1021/ac800112r]
4. Katuwal G, Pandey S, Hennessey M, Lamichhane B. Applications of blockchain in healthcare: current landscape & challenges. arXiv. Preprint posted online on December 6, 2018. [FREE Full text]
5. Siyal AA, Junejo AZ, Zawish M, Ahmed K, Khalil A, Soursou G. Applications of blockchain technology in medicine and healthcare: challenges and future perspectives. Cryptography 2019 Jan;3(1):3. [doi: 10.3390/cryptography3010003]
6. Liu H, Crespo RG, Martínez OS. Enhancing privacy and data security across healthcare applications using blockchain and distributed ledger concepts. Healthcare 2020 Jul;8(3):243 [FREE Full text] [doi: 10.3390/healthcare8030243] [Medline: 32751325]
7. Bouras MA, Lu Q, Zhang F, Wan Y, Zhang T, Ning H. Distributed ledger technology for eHealth identity privacy: state of the art and future perspective. Sensors 2020 Jan;20(2):483. [doi: 10.3390/s20020483]
8. Zheng X, Sun S, Mukkamala RR, Vatrapu R, Ordieres-Meré J. Accelerating health data sharing: a solution based on the internet of things and distributed ledger technologies. J Med Internet Res 2019 Jun;21(6):e13583 [FREE Full text] [doi: 10.2196/13583] [Medline: 31172963]
9. Brogan J, Baskaran I, Ramachandran N. Authenticating health activity data using distributed ledger technologies. Comput Struct Biotechnol J 2018;16:257-266 [FREE Full text] [doi: 10.1016/j.csbj.2018.06.004] [Medline: 30101004]
10. Tasatanattakool P, Techapanupreeda C. Blockchain: Challenges and applications. : IEEE Computer Society; 2018 Oct 18 Presented at: International Conference on Information Networking; October 18, 2018; Bangkok p. 473-475. [doi: 10.1109/ICOIN.2018.8343163]
11. Erokhin A, Koshechkin K, Ryabkov I. The distributed ledger technology as a measure to minimize risks of poor-quality pharmaceuticals circulation. PeerJ Comput Sci 2020;6:e292 [FREE Full text] [doi: 10.7717/peerj-cs.292] [Medline: 33816943]
12. Kordestani H, Barkaoui K, Zahran W. HapiChain: a blockchain-based Framework for patient-centric telemedicine. : Institute of Electrical and Electronics Engineers; 2020 Presented at: IEEE 8th International Conference on Serious Games and Applications for Health (SeGAH); August 12-14, 2020; Vancouver p. 1-6. [doi: 10.1109/SeGAH49190.2020.9201726]
13. Griggs KN, Ossipova O, Kohlios CP, Baccarini AN, Howson EA, Hayajneh T. Healthcare blockchain system using smart contracts for secure automated remote patient monitoring. J Med Syst 2018 Jun;42:130. [doi: 10.1007/s10916-018-0982-x]
14. Toshtemirov A, Mihuba G, Shundi MJ. Using blockchain technology for design of telemedicine system based on mobile terminal. Int J Emerg Technol Eng Res 2018;6(12):12-19.
15. Colón KA. Creating a patient-centered, global, decentralized health system: combining new payment and care delivery models with telemedicine, AI, and blockchain technology. BHTY 2018 Sep;1:1-18. [doi: 10.30953/bhty.v1.30]
16. Mackey T, Bekki H, Matsuzaki T, Mizushima H. Examining the potential of blockchain technology to meet the needs of 21st-nentury Japanese health care: viewpoint on use cases and policy. J Med Internet Res 2020 Jan;22(1):e13649 [FREE Full text] [doi: 10.2196/13649] [Medline: 31917371]
17. Nusrat SA, Ferdous J, Ajmat SA, Ali A, Sorwar G. Telemedicine system design using blockchain in Bangladesh. 2020 Presented at: IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE); December 9-11, 2019; Melbourne p. 1-5. [doi: 10.1109/CSDE48274.2019.9162401]
18. Aileni RM, Suciu G. IoMT: a blockchain perspective. In: Khan M, Quasim M, Algarni F, Alharthi A, editors. Decentralised Internet of Things. Switzerland: Springer, Cham; 2020:199-215.
19. Angraal S, Krumholz HM, Schulz WL. Blockchain technology. Circ Cardiovasc Qual Outcomes 2017 Sep;10(9):1. [doi: 10.1161/circoutcomes.117.003800]
20. Kuo T, Kim HE, Ohno-Machado L. Blockchain distributed ledger technologies for biomedical and health care applications. J Am Med Inform Assoc 2017 Nov;24(6):1211-1220 [FREE Full text] [doi: 10.1093/jamia/ocx068] [Medline: 29016974]
21. Mettler M. Blockchain technology in healthcare: the revolution starts here. 2016 Presented at: IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom); September 14-16, 2016; Munich p. 1-3. [doi: 10.1109/healthcom.2016.7749510]
22. Ahmad RW, Salah K, Jayaraman R, Yaqoob I, Ellahham S, Omar M. The role of blockchain technology in telehealth and telemedicine. Int J Med Inform 2021 Apr;148:104399 [FREE Full text] [doi: 10.1016/j.ijmedinf.2021.104399] [Medline: 33540131]
23. Azim A, Islam MN, Spranger PE. Blockchainnovel coronavirus: towards preventing COVID-19 future pandemics. Iberoam J Med 2020;2(3):215-218. [doi: 10.5281/zenodo.3779244]

24. Monaghesh E, Hajizadeh A. The role of telehealth during COVID-19 outbreak: a systematic review based on current evidence. BMC Public Health 2020 Aug;20(1):1193 [FREE Full text] [doi: 10.1186/s12889-020-09301-4] [Medline: 32738884]
25. Telehealth Continues To Change The Face Of Healthcare Delivery - For The Better Electronic Resource. URL: https://www.forbes.com/sites/jeffgorke/2019/09/24/telehealth-continues-to-change-the-face-of-healthcare-delivery-for-the-better/?sh=74889a89565f [accessed 2021-04-01]
26. Azaria A, Ekblaw A, Vieria T, Lippman A. MedRec: using blockchain for medical data access and permission management. 2016 Presented at: 2016 2nd International Conference on Open and Big Data (OBD); August 22-24, 2016; Vienna p. 25-30. [doi: 10.1109/OBD.2016.11]
27. Armstrong S. Bitcoin technology could take a bite out of NHS data problem. BMJ 2018 May:k1996. [doi: 10.1136/bmj.k1996]
28. Symptomatic - Timelines. URL: https://www.symptomatic.io/marketing-page [accessed 2021-04-01]
29. What we believe: the safety of the people shall be the highest law. Marcus Tullius Cicero Quotes. URL: https://quotepark.com/authors/marcus-tullius-cicero/ [accessed 2021-04-01]
30. Robomed Network Launches Blockchain-Based Health Services Platform. URL: https://www.itweek.ru/blockchain/news-company/detail.php?ID=204320 [accessed 2021-04-01]
31. Healing Technologies: How Does the Big Data Help Physicians to Treat Patients. URL: https://medium.com/robomed [accessed 2021-04-01]
32. Business Process Management. URL: https://robo-med.com/about/ [accessed 2021-04-01]
33. Mannaro K, Baralla G, Pinna A, Ibba S. A blockchain approach applied to a teledermatology platform in the Sardinian region (Italy). Information 2018 Feb;9(2):44. [doi: 10.3390/info9020044]
34. A New Paradigm in Healthcare Data Privacy AND Security. URL: http://www.myhealthmydata.eu/ [accessed 2021-04-01]
35. Jarzabek G, Schucht P, Rzeczkowski L. Patient and doctor-centric health ecosystem in specialised medicine engineered on blockchain. TrustedHealth 2017:33-34 [FREE Full text] [doi: 10.30748/soi.2021.164.04]
36. Skychain Global - Blockchain Infrastructure for Medical AI. URL: https://skychain.global/wp-content/uploads/2018/03/wp_english-Newest.pdf [accessed 2021-04-01]
37. Merkow J, Lufkin R, Nguyen K, Soatto S, Tu Z, Vedaldi A. DeepRadiologyNet: radiologist level pathology detection in CT head images. ArXiv. Preprint posted online on November 26, 2017. [FREE Full text]
38. eHealth First. URL: https://ehfirst.io/ [accessed 2021-04-01]
39. A new healthcare payment method using tokens on a blockchain. URL: https://neironix.io/documents/whitepaper/3185/whitepaper.pdf [accessed 2021-04-01]
40. Yusuf H, Bergström MM. Creating legitimacy within blockchain startups in a virtual context - A study on how decentralized organisations raise funds before executing an Initial Coin Offering [Dissertation].: The Institution for Informatics and Media, Uppsala University; 2018. URL: https://www.diva-portal.org/smash/get/diva2:1286073/FULLTEXT01.pdf [accessed 2021-04-01]
41. HealPoint | AngelList Talent. URL: https://angel.co/company/healpoint [accessed 2021-04-01]
42. Elcoin. URL: https://elcoin.io/White_Paper%200.9_v3-final.pdf [accessed 2021-04-01]
43. Atutov SN, Calabrese R, Plekhanov AI, Tomassetti L. Diffusion and photodesorption of molecular gases in a polymer organic film. Eur Phys J D 2014 Jan;68(1):1. [doi: 10.1140/epjd/e2013-40468-7]
44. PointNurse Reaches Milestone on Path to Digital Health Provider Leadership. URL: https://www.prnewswire.com/news-releases/pointnurse-reaches-milestone-on-path-to-digital-health-provider-leadership-300974449.html [accessed 2021-04-01]
45. Engelhardt MA. Hitching healthcare to the chain: an introduction to blockchain technology in the healthcare sector. TIM Review 2017 Oct;7(10):22-34. [doi: 10.22215/timreview/1111]
46. Todaro J. The community asked us to build. And that’s what we did. Preparing for the Launch of MedX. URL: https://medium.com/medxprotocol/preparing-for-the-launch-of-medx-97d39e1d9e9c [accessed 2021-04-01]
47. Wong DR, Bhattacharya S, Butte AJ. Prototype of running clinical trials in an untrustworthy environment using blockchain. Nat Commun 2019 Feb;10(1):917 [FREE Full text] [doi: 10.1038/s41467-019-08874-y] [Medline: 30796226]
48. Distributed Ledger in Healthcare: Why Medicine Needs Blockchain Electronic Resource. URL: https://tallinn.mhealth.events/en/article/raspredelenniy-reestr-v-oblasti-zdravoohraneniya-pochemu-meditsine-nugen-blokcheyn-96363 [accessed 2021-04-01]

Abbreviations

AI: artificial intelligence
EHR: electronic health record

Edited by G Eysenbach; submitted 15.12.19; peer-reviewed by F Lehocki, JT te Gussinklo, S Pit, E Chukwu, A Kamisalic, L Luo, C Reis, J Liu, R Hylock, TT Kuo; comments to author 26.10.20; revised version received 12.02.21; accepted 09.05.21; published 18.08.21.

Please cite as:
Koshechkin K, Lebedev G, Radzievsky G, Seepold R, Martinez NM
Blockchain Technology Projects to Provide Telemedical Services: Systematic Review
J Med Internet Res 2021;23(8):e17475
URL: https://www.jmir.org/2021/8/e17475
doi:10.2196/17475
PMID:34407924

©Konstantin Koshechkin, Georgy Lebedev, George Radzievsky, Ralf Seepold, Natividad Madrid Martinez. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 18.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Chinese Patients' Intention to Use Different Types of Internet Hospitals: Cross-sectional Study on Virtual Visits

Liyun Liu[1], PhD; Lizheng Shi[2], PhD

[1]School of Humanities and Management, Zhejiang Chinese Medical University, Hangzhou, China
[2]School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA, United States

Corresponding Author:
Liyun Liu, PhD
School of Humanities and Management
Zhejiang Chinese Medical University
No. 260 Baichuan Street
Fuyang District
Hangzhou, 311402
China
Phone: 86 0571 8661 3687
Email: llyun@zcmu.edu.cn

Abstract

Background: The issuing of regulation schemes and the expanding health insurance coverage for virtual visits of internet hospitals would incentivize Chinese providers and patients to use virtual visits tremendously. China's internet hospitals vary in sponsorship. However, little is known about patients' intention to use virtual visits delivered by different sponsorship types of internet hospitals.

Objective: The goal of the research is to examine patients' intention to use virtual visits, as well as virtual visits delivered by different sponsorship types of internet hospitals. In addition, we will identify determinants of patients' intention to use virtual visits, as well as intention to use virtual visits delivered by different sponsorship types of internet hospitals.

Methods: A cross-sectional survey of 1653 participants was conducted in 3-tier hospitals in 3 cities with different income levels in May and June 2019. Binary logistic regression analysis was used to identify the factors that affect patients' intention to use virtual visits. Multinomial logistic regression analysis was conducted to identify the determinants of the intention to use virtual visits delivered by different sponsorship types of internet hospitals (ie, enterprise-sponsored, hospital-sponsored, and government-sponsored).

Results: A total of 76.64% (1145/1494) of adult participants were online medical information seekers, and 87.06% (969/1113) of online medical information seekers had intention to use virtual visits. Public hospital–sponsored internet hospitals were the most prevalent ones among Chinese patients (473/894, 52.9%), followed by the provincial government internet hospital platform (238/894, 26.6%), digital health companies (116/894, 13.0%), medical e-commerce companies (48/894, 5.4%), private hospitals (13/894, 1.5%), and other companies (6/894, 0.7%). Gender, education, monthly income, and consumer type were significantly associated with the intention to use virtual visits. Gender, age, education, city income level, consumer type, and trust in the sponsor of a health website were significantly associated with the patient's intention to use virtual visits delivered by 3 different sponsorship types of internet hospitals.

Conclusions: Chinese patients who were online medical information seekers had high intention to use virtual visits and had different intentions to use virtual visits delivered by different sponsorship types of internet hospitals. Public hospitals, the government, and digital health companies were the top 3 sponsorship types of internet hospitals that patients had intention to use. Trust in a health website sponsor significantly influenced the patient's intention to use virtual visits delivered by different sponsorship types of internet hospitals. Gender, education, and consumer type were the factors significantly associated with both the intention to use virtual visits and the intention to use virtual visits delivered by different sponsorship types of internet hospitals.

(J Med Internet Res 2021;23(8):e25978) doi:10.2196/25978

KEYWORDS

internet hospital; direct-to-consumer telemedicine; virtual visit; trust; intention to use; sponsorship type

Introduction

Background

Virtual consultation and virtual visit are two primary types of services delivered by internet hospitals in China. An internet hospital is, to a large extent, an equivalent of direct-to-consumer telemedicine. China's first internet hospital officially opened in 2014 [1,2]; however, since then the internet hospital industry witnessed an initial development stage with ups and downs. Virtual consultation was the primary service type of internet hospitals from 2014 to 2018. The difference between virtual consultation and virtual visit primarily resulted from the evolution of China's internet hospital regulation schemes. The State Council of China issued the guideline on Internet Plus Healthcare of 2018 [3], and the National Health Commission accordingly issued specific regulation schemes for online medical diagnosis and treatment as well as internet hospitals in 2018 [4]. The issuing of regulation schemes have brought the rapid development of internet hospitals [5], especially for the service type of online medical diagnosis and treatment. Online medical diagnosis and treatment (hereinafter referred to as virtual visit) is very similar to a virtual visit of direct-to-consumer telemedicine/telehealth in many other countries like the United States because physicians are able to diagnose, treat, and prescribe for some common conditions and chronic diseases for non–first-visit patients. However, physicians are not allowed to diagnose, treat, and prescribe in the virtual consultation service.

In addition to the difference in scope of service, virtual visit is also different from consultation in health insurance coverage. Previous research has suggested that health insurance coverage is a significant factor that affects the use of virtual visits [6]. Following the guideline on Internet Plus Healthcare, in August 2019, the National Healthcare Security Administration of China, as the single payer, has accordingly issued specific guidelines aiming to expand insurance coverage for virtual visit services of internet hospitals [7]. However, consultation services of internet hospitals have not been covered by health insurance. Obviously, health insurance coverage would incentivize both providers and consumers to use virtual visits tremendously. However, there is a lack of studies that specifically examine consumers' intention to use virtual visits in China.

Internet hospitals vary in sponsors. There are primarily brick-and-mortar hospital-sponsored and enterprise-sponsored internet hospitals in China's direct-to-consumer telemedicine market [5]. Recently, a very small number of local governments have taken initiatives to set up local government internet hospital platforms by pooling local public health care resources. Hence, enterprises, hospitals and governments are 3 major sponsor types of internet hospitals. Han et al [8] demonstrated that different initiators/sponsors of internet hospitals including the government, hospitals, and enterprises have different purposes and scopes of service (ie, target consumer). The virtual visit market is an emerging market in China, and hence a study on consumers' intention to use virtual visits delivered by different sponsorship types of internet hospitals is critically important to understand the market structure and future development of the industry. However, little is known about consumers' intention to use different sponsorship types of internet hospitals in China or intention to use virtual visits delivered by different sponsorship types of internet hospitals.

Study Aims

We conducted a cross-sectional survey in Zhejiang province to examine patients' intention to use virtual visits and their intention to use virtual visits delivered by different sponsorship types of internet hospitals and identify the factors that affect patients' intention to use virtual visits. Zhejiang takes a leading role in China's internet hospital industry development because it has the first licensed enterprise-sponsored internet hospital (WeDoctor Group) [9] and the first public tertiary hospital–sponsored one (the First Affiliated Hospital of Zhejiang University) [10], the first direct-to-consumer provincial government internet hospital platform [11].

When our survey was conducted, Zhejiang was the first and only province in China that delivered direct-to-consumer telemedicine services including virtual visits for its residents via its provincial internet hospital platform. Only after the outbreak of COVID-19 did other provincial governments start to deliver direct-to-consumer telemedicine service to their residents; however, the majority of services were specially designed free virtual consultations to contain the COVID-19 epidemic. Zhejiang is currently still the leading province to deliver virtual visits by pooling all local health care resources via its provincial internet hospital platform. Therefore, a survey on residents in Zhejiang Province can provide prospective insights for research questions on Chinese consumers' intention to use virtual visits delivered by hospital-sponsored, enterprise-sponsored, and government-sponsored internet hospitals.

Literature Review and Hypotheses

As two primary service types of internet hospitals, virtual visit and virtual consultation are different from each other in service scope and health insurance coverage. However, there is hardly a clear division between virtual consultation and virtual visit in the current knowledge of internet hospital literature. Previous studies have examined either patients' intention to use an internet hospital [12,13] or a virtual consultation delivered by internet hospitals [14]. As noted in the introduction, a virtual visit in China is equivalent to a direct-to-consumer telemedicine/telehealth visit in many other countries. However, most previous studies in the literature have examined patients' intention to use or actual use of direct-to-consumer telemedicine from the perspectives of specific disease-related application [15,16] and specific telemedicine website/app [17,18]. There is a lack of studies on patients' decision on sponsorship types of internet hospital platforms and determinants of patients' decision making.

Systematic reviews have revealed that the technology acceptance model (TAM) was the most commonly used model to examine users' acceptance of health information technologies including telemedicine [19-21]. The TAM beliefs (technological issues: perceived ease of use, perceived usefulness) and consumer trust are two distinct sets of beliefs that contribute in their own right to increase intention to use the website and, through it, transactions with the e-vendor [22]. Through a systematic review and meta-analysis, Tao et al [19] found that trust is significantly correlated with behavioral intention to use consumer-oriented health information technologies including telemedicine. Trust positively influences the patient's intention to use telemedicine services in developing countries like Pakistan [23]. However, very few previous studies have examined the patient's trust in a telemedicine service including trust in the care organization, trust in the care professional, trust in the treatment, and trust in the technology [24]. There is a lack of studies that examine patients' trust in the sponsor/owner. Based on the abovementioned findings, we hypothesized that trust in the sponsor significantly affected Chinese patients' intention to use virtual visits delivered by different sponsorship types of internet hospitals (Hypothesis 1).

Websites, apps and WeChat public accounts are 3 primary modalities through which Chinese providers deliver direct-to-consumer online health information and services. Health websites and apps in China usually cover provisions of health/medical information and knowledge, in-person visit online appointment, result tracing of laboratory and diagnostic imaging tests, and consultation, etc. For some health websites that have been licensed to deliver internet hospital services, the internet hospital, especially a virtual visit, is also an essential part of these health websites. However, among all types of online health information services, Chinese patients had the highest awareness and use of in-person visit appointments and medical fee payment online [25]. Therefore, internet hospital service is an important part of health websites, but it is still not an essential part of many health websites in China, especially before the outbreak of COVID-19. Accordingly, most Chinese patients only had prior knowledge or experience of health websites, which led to the result that they formed perceived trust in health websites rather than internet hospital platforms.

The classification of health websites is also different from that of internet hospitals in the current literature. The website owner/sponsor has been identified as one of the most widely reported indicators consumers applied to evaluate the quality of online health information [26,27]. Internet users' perceived trust in online health information and service delivered by the health website varies by its sponsor/owner type. Health websites vary in their content and features, across commercial, governmental, and nonprofit websites, as they must respond to different structural incentives and constraints, motivations, and purposes [28,29]. Although there are some differences in sponsorship categories of health websites in previous studies, health websites are usually categorized as commercial, governmental, organizational (eg, nonprofit medical institutions), educational (universities and academic institutions), and personal [29-31]. Users usually perceived medical institutions, universities, and governments as the more trustworthy health website sponsors/owners [32]. Based on the findings that users' trust in health websites varied by the website sponsorship type and trust affected behavioral intention, we concluded that patients' trust in health website sponsor affected their intention to use health websites and patients' intention to use a health website varied by its sponsorship type. Based on the fact that an internet hospital was an important part of a health website, we hypothesized that patients' intention to use virtual visits varied by sponsorship types of internet hospitals (Hypothesis 2).

Methods

With reference to the Health Information National Trends Survey (HINTS) of the US National Cancer Institute, we made some modifications in accordance with China's context and research questions to design our survey questionnaire (Multimedia Appendix 1).

Participants and Data Collection

We conducted a cross-sectional survey using stratified sampling in 3 cities of different income levels in Zhejiang Province, China. According to Zhejiang Statistical Yearbook 2019 [33], 5 cities were categorized as high-income cities (per capita gross domestic product [GDP] > US $15,000), 3 cities were categorized as medium-income cities (US $15,000 ≥ per capita GDP ≥ US $10,000), and 3 cities were categorized as low-income cities (per capita GDP < US $10,000). Proportionate sampling was used to determine the weight of each income-level city stratum. The weight of the stratum was the proportion of the population contained in that stratum, and the population was the urban population of Zhejiang Province in 2018 [33]. The sampling weight of high-income cities was 52.72%, that of medium-income cities was 24.24%, and that of low-income cities was 23.04%.

A pretest with a sample of 39 participants was conducted in April 2019. Based on China's 3-tier health care delivery system, hospitals are designated as primary, secondary (secondary 2A, 2B), and tertiary institutions (tertiary 3A, 3B). This study conducted an in-person structured questionnaire survey in 5 different types of hospitals (tertiary 3A, 3B, secondary 2A, 2B, and primary) in May and June 2019. A total of 1653 participants were then randomly surveyed in the outpatient center of each hospital, and health practitioners were excluded from the survey. Hangzhou was the high-income city (n=871), Jinhua was the medium-income city (n=401), and Lishui was the low-income level city (n=381).

Measure

Sociodemographic Characteristics

Sociodemographic characteristics of gender, age, education level, and marital status were included in the analysis. Monthly income level (low-income: less than CNY 1800 [less than US $272]; medium-low income: CNY 1801-4600 [US $272-$696]; medium-income: CNY 4601-8000 [US $697-$1210]; medium-high-income: CNY 8001-17,000 [US $1211-$2571]; and high-income: more than CNY 17,000 [more than US $2571]) was included. Participants were asked to rate their current health status using a 5-point Likert scale ranging from very good (1) to very poor (5).

Internet Use and Medical Information Seeking

According to Internet Development Report of Zhejiang Province 2019, 80.9% of its residents had access to the internet [34]. Therefore, we modified the HINTS question measuring internet use by asking: “Is the internet your major source of information?” Those participants answering yes were subsequently referred to as active internet users. We modified HINTS questions concerning medical/health information seeking by asking: “Have you used the internet to look for medical information before?” Those who answered yes were subsequently referred to as online medical information seekers; those answered no were excluded from the final analysis.

Trust in Health Website Sponsor

We measured the trust in health website sponsorship types among Chinese patients by asking: “Which type of health website do you perceive as the most trustworthy source of medical information?” Websites of medical schools/universities and academic institutions do not provide direct-to-consumer online medical information and services for lay persons; consequently, there are no educational health websites in China. Based on China’s context, we categorized health websites into 4 website sponsorship types: governmental, commercial (digital health companies and internet companies, etc), organizational (hospitals), and personal (individual health practitioners). The option of other was included in case there were some participants who had no prior experience and knowledge of health websites. Rice et al [28] referred to health websites of nonprofit sectors as organizational. In terms of capacities, the majority of hospitals in China are nonprofit public hospitals [35], and there are few other nonprofit organizations providing online direct-to-consumer medical/health information for lay persons. Therefore, the organizational health website thereafter denoted public hospitals.

Intention to Use Virtual Visits

We measured the intention to use virtual visits of internet hospitals by asking: “The government has issued guidelines on the development of internet hospitals. Virtual visits of internet hospitals enable physicians to diagnose, treat, and prescribe for some common conditions and chronic diseases for non–first-visits and deliver the prescription right to your door. Do you have the intention to use the virtual visit?”

Intention to Use Virtual Visits Delivered by Different Sponsorship Types of Internet Hospitals

We measured the intention to use virtual visits delivered by different sponsorship types of internet hospitals by asking: “Which type of internet hospital do you have the highest intention to use regarding a virtual visit?” There were the facts regarding sponsorship types of internet hospitals in China: (1) hospital sponsors included public hospitals and private hospitals; enterprise sponsors included digital health companies, internet tech companies, medical e-commerce companies, medical informatics companies, health management companies, hospital management companies, pharmaceutical companies, insurance companies, medical equipment companies, etc [5] and (2) the majority of sponsors of enterprise-sponsored internet hospitals were digital health companies and medical e-commerce companies by June 2019 [36]. Therefore, sponsorship types of internet hospitals included 6 options: digital health companies (eg, WeDoctor, Haodf); medical e-commerce companies (eg, Ali Health, JD Health); other companies (enterprise sponsors except digital health companies and medical e-commerce companies); public hospitals; private hospitals; and the Zhejiang Provincial Internet Hospitals Platform.

Statistical Analyses Strategy

Stata 12.0 software (StataCorp LLC) was used to conduct statistical analyses. Descriptive statistics identified sociodemographic characteristics of the final sample, proportions of participants who had the intention to use virtual visits, and proportions of participants who had the intention to use virtual visits delivered by different sponsorship types of internet hospitals.

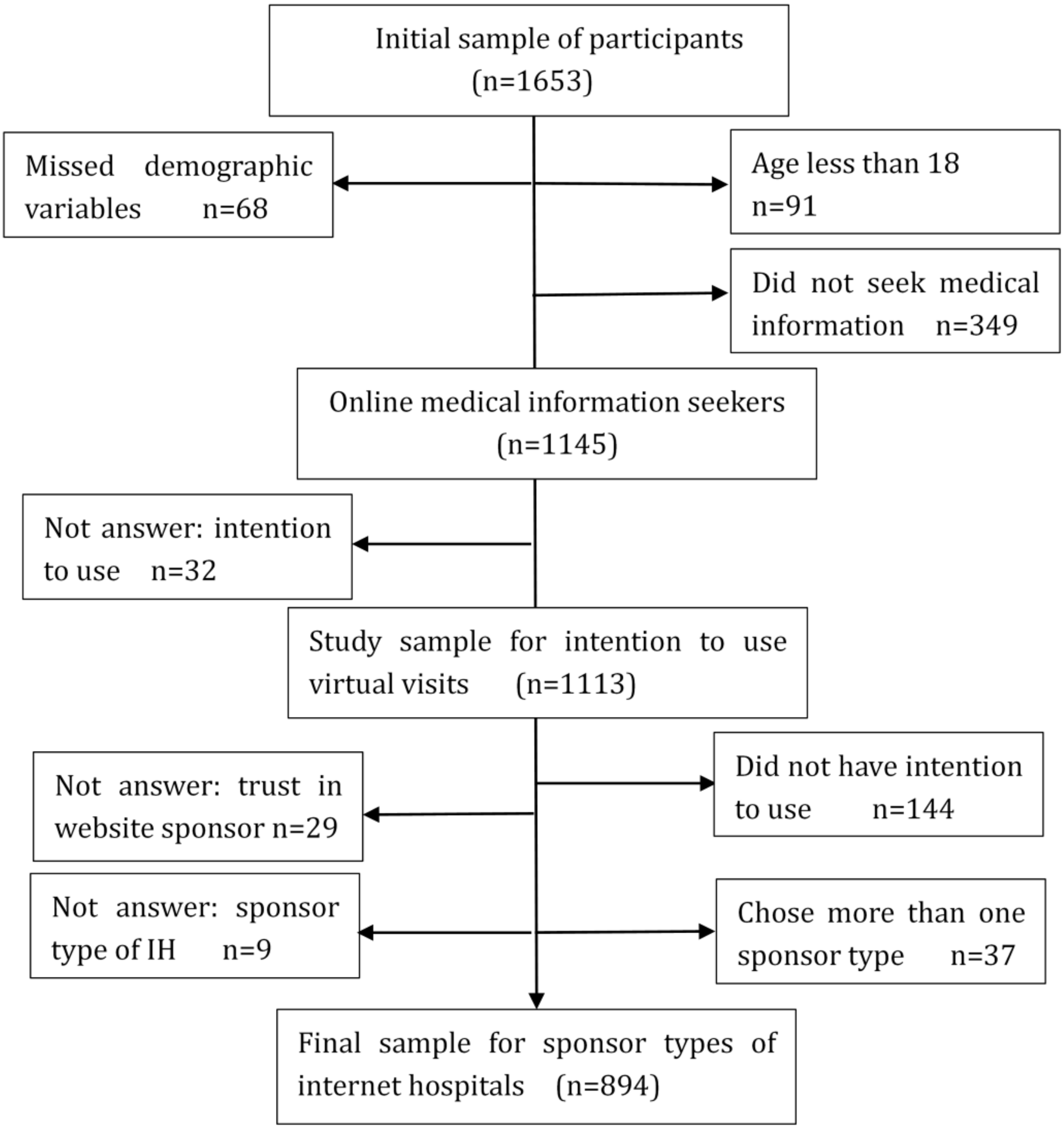
A systematic review of end user acceptance of telemedicine use [21] found that logistic regression, structural equation modeling, and linear regression were 3 primary statistical analysis methods conducted in previous studies. It also found that TAM was the most used model. Structural equation modeling was primarily used to examine the relationships between factors like TAM beliefs (perceived ease of use, perceived usefulness) and behavioral intention to use telemedicine. And there were usually several paths through which different factors affected the behavioral intention. However, in this study, there was only one path (ie, the trust in sponsor affected patients’ intention to use virtual visits delivered by different types of internet hospitals). Additionally, the sponsorship type of internet hospital was a qualitative variable (categorical variable). Therefore, logistic regression was the appropriate statistical technique to identify the determinants. Binary logistic regression analysis was used to identify the factors that affect patients’ intention to use the virtual visit. Multinomial logistic regression analysis was conducted to examine the association of patient characteristics with the intention to use virtual visits delivered by different sponsorship types of internet hospitals.

Results

Sample Selection

Health information seeking was the patient’s most frequently used eHealth or mobile health (mHealth) activity [37], and health information seeking was also a significant predictor of eHealth and mHealth service use [38,39], therefore we inferred that online medical information seekers were target consumers for internet hospitals. To identify the target market, 349 participants were excluded because they self-reported that they have not sought medical information online before. The study sample for statistical analysis on intention to use virtual visits was 1113. The final sample for statistical analysis on intention to use virtual visits delivered by different sponsorship types of internet hospitals was 894. Figure 1 shows the flowchart of sample selection.

Figure 1. Flowchart of sample selection. IH: internet hospital.


Initial sample of participants (n=1653)
Missed demographic variables n=68
Age less than 18 n=91
Did not seek medical information n=349
Online medical information seekers (n=1145)
Not answer: intention to use n=32
Study sample for intention to use virtual visits (n=1113)
Not answer: trust in website sponsor n=29
Did not have intention to use n=144
Not answer: sponsor type of IH n=9
Chose more than one sponsor type n=37
Final sample for sponsor types of internet hospitals (n=894)

Sociodemographic Characteristics

Among adult participants, 76.64% (1145/1494) were online medical information seekers. Among medical information seekers, those who did not answer the question of intention to use were excluded. Table 1 shows the sociodemographic characteristics of the final sample for intention to use virtual visits. There were more female medical information seekers than male seekers, which was possibly because females were more likely to be caregivers and information seekers who accompanied family members to visit hospitals. The majority (767/1113, 68.91%) of seekers had the education level of college graduate. Young seekers aged 18 to 55 years accounted for 98.11% (1092/1113) of participants. The survey was conducted in cities in Zhejiang Province where cities usually had both high stocks and inflows of well-educated young people. The majority (965/1113, 86.71%) of seekers were in good or fair health status. Compared to other variables, monthly income was more equally distributed. A total of 80.41% (895/1113) were active internet users whose major information source was the internet. A total of 87.06% (969/1113) reported that they had intention to use virtual visits of internet hospitals.

Table 1. The sociodemographic characteristics of study sample for intention to use virtual visits (n=1113).

Characteristic	Value, n (%)
Gender	
Male	445 (39.98)
Female	668 (60.02)
Age (years)	
18-29	508 (45.64)
30-40	430 (38.63)
41-55	154 (13.84)
56-65	15 (1.35)
>65	6 (0.54)
Health status	
Very good	110 (9.88)
Good	482 (43.31)
Fair	483 (43.40)
Poor	36 (3.23)
Very poor	2 (0.18)
Marital status	
Married	658 (59.12)
Single	434 (38.99)
Divorced	19 (1.71)
Widowed	2 (0.18)
Education level	
≤Junior high school	101 (9.07)
Senior high school	166 (14.91)
College graduate	767 (68.91)
Postgraduate	79 (7.11)
Monthly income (CNY)	
≤1800	122 (10.96)
1801-4600	326 (29.29)
4601-8000	350 (31.45)
8001-17,000	226 (20.30)
>17,000	89 (8.00)
Internet use	
Inactive user	218 (19.59)
Active user	895 (80.41)
Intention to use virtual visits	
No	144 (12.94)
Yes	969 (87.06)

Intention to Use Virtual Visits

The binary logistic regression result of intention to use virtual visits of internet hospitals is presented in Table 2. The female was more likely to have intention to use virtual visits (odds ratio [OR] 1.68, 95% CI 1.13-2.48). Education level was significantly and positively associated with the intention to use virtual visits, senior high school (OR 2.05, 95% CI 1.09-3.86), college graduate (OR 3.44, 95% CI 1.94-6.12), postgraduate (OR 3.09, 95% CI 1.19-8.00). Patients with high income level were more likely than those with low income level to have intention to use virtual visits (OR 3.22, 95% CI 1.13-9.15). Consumers of tertiary B hospitals were more likely than consumers of primary hospitals to have intention to use virtual visits (OR 2.62, 95% CI 1.12-6.13).

Table 2. Binary logistic regression result of intention to use virtual visits of internet hospitals (n=1113).

Characteristic	OR[a] (95% CI)	*P* value
Gender (ref[b]: Male)	1.68 (1.13-2.48)	.01
Age (ref: 18-29)		
30-40	0.92 (0.54-1.58)	.77
41-55	1.18 (0.58-2.38)	.65
56-65	0.62 (0.15-2.53)	.51
>65	0.71 (0.11-4.65)	.72
Health status	1.14 (0.89-1.47)	.29
Marital status (ref: Married)		
Single	0.75 (0.44-1.30)	.31
Divorced	0.69 (0.20-2.33)	.55
Widowed	—[c]	—
Education level (ref: ≤Junior high school)		
Senior high school	2.05 (1.09-3.86)	.03
College graduate	3.44 (1.94-6.12)	<.001
Postgraduate	3.09 (1.19-8.00)	.02
Monthly income (CNY; ref: ≤1800)		
1801-4600	1.00 (0.54-1.83)	.99
4601-8000	1.67 (0.88-3.17)	.12
8001-17,000	1.51 (0.75-3.03)	.25
>17,000	3.22 (1.13-9.15)	.03
Internet use (ref: Inactive)	1.37 (0.88-2.13)	.17
City (ref: Hangzhou)		
Jinhua	1.37 (0.80-2.35)	.25
Lishui	1.26 (0.73-2.18)	.41
Hospital type (ref: Primary)		
Secondary B	0.64 (0.23-1.82)	.41
Secondary A	1.89 (0.71-5.09)	.21
Tertiary B	2.62 (1.12-6.13)	.03
Tertiary A	1.95 (0.94-4.06)	.07
Constant	0.45 (0.13-1.54)	.20

[a] OR: odds ratio.

[b] ref: reference group.

[c] Not applicable.

Intention to Use Virtual Visits Delivered by Different Types of Internet Hospitals

Table 3 presents Chinese patients' intention to use virtual visits delivered by different types of internet hospitals. Public hospital–sponsored internet hospitals were the most prevalent (473/894, 52.9%), followed by the provincial government internet hospital platform (238/894, 26.6%), digital health companies (116/894, 13.0%), medical e-commerce companies (48/894, 5.4%), private hospitals (13/894, 1.5%), and other companies (6/894, 0.7%). This result was consistent with the current market structure of internet hospitals in China. Brick-and-mortar hospital sponsors and enterprise sponsors were 2 primary sponsor types of internet hospitals in China, while hospital-sponsored ones accounted for 83.5% (415/497) of internet hospitals by April 30, 2020. Within the category of hospital-sponsored internet hospitals, public hospitals and private hospitals accounted for 90.4% (375/415) and 9.6% (40/415) [5].

Multinomial logistic regression was used to identify the determinants of intention to use virtual visits delivered by 3 major different sponsorship types of internet hospitals (enterprise-sponsored, hospital-sponsored, and government-sponsored). The final sample for the multinomial logistic regression analysis was 894. Table 4 shows the result of multinomial logistic regression analysis.

Using hospital-sponsored internet hospitals as the base outcome, we had following results. Females were more likely to have intention to use virtual visits delivered by the governmental internet hospital platform and less likely to have intention to use visits delivered by enterprise-sponsored internet hospitals. Participants aged 30 to 40 years were more likely than those aged 18 to 29 years to have intention to use the governmental internet hospital platform. Participants with higher education levels were less likely to have intention to use virtual visits delivered by enterprise-sponsored internet hospitals. Participants in a medium-income city (Jinhua) were less likely than those in a high-income city (Hangzhou) to have intention to use virtual visits delivered by both enterprise-sponsored internet hospitals and governmental internet hospital platform. Consumers of secondary A, tertiary B, and tertiary A hospitals were less likely than consumers of primary hospitals to have intention to use enterprise-sponsored internet hospitals. Participants who trusted commercial health websites most were more likely than those trusted governmental health websites to have intention to use virtual visits delivered by enterprise-sponsored internet hospitals rather than visits delivered by hospital-sponsored ones. Compared to participants who trusted in governmental health websites, participants who trusted all other sponsorship types of health websites were less likely to use virtual visits delivered by the governmental internet hospital platform.

Table 3. Patients' intention to use virtual visits delivered by different types of internet hospitals (n=894).

Internet hospital type	Value, n (%)
Enterprise-sponsored	170 (19.0)
Digital health companies	116 (13.0)
Medical e-commerce companies	48 (5.4)
Other companies[a]	6 (0.7)
Hospital-sponsored	486 (54.4)
Public hospitals	473 (52.9)
Private hospitals	13 (1.5)
Government-sponsored	238 (26.6)

[a] Other companies: all other enterprise sponsors except digital health companies and medical e-commerce companies.

Table 4. Multinomial logistic regression result of intention to use virtual visits delivered by different sponsorship types of internet hospitals (base outcome=hospital; n=894).

Characteristic	Enterprise		Government	
	RRR[a] (95% CI)	*P* value	RRR (95% CI)	*P* value
Gender (ref[b]: Male)	0.66 (0.44-0.98)	.04	1.57 (1.09-2.28)	.02
Age (years; ref: 18-29)				
30-40	1.01 (0.59-1.74)	.96	1.74 (1.07-2.83)	.03
41-55	0.75 (0.36-1.56)	.44	0.83 (0.42-1.63)	.59
56-65	1.85 (0.36-9.36)	.46	0.54 (0.05-6.19)	.62
>65	1.39 (0.10-18.34)	.80	—[c]	.99
Health status	0.86 (0.66-1.12)	.25	1.04 (0.82-1.32)	.73
Marital status (ref: Married)				
Single	1.36 (0.80-2.33)	.25	1.40 (0.86-2.27)	.18
Divorced	0.50 (0.06-4.25)	.52	1.19 (0.36-3.91)	.78
Widowed	—	.99	—	.99
Education (ref: ≤Junior high school)				
Senior high school	0.33 (0.14-0.77)	.01	0.67 (0.29-1.53)	.34
College graduate	0.46 (0.22-0.96)	.04	0.97 (0.45-2.09)	.94
Postgraduate	0.20 (0.06-0.61)	.01	0.73 (0.27-1.98)	.54
Monthly income (CNY; ref: ≤1800)				
1801-4600	0.83 (0.42-1.63)	.58	0.72 (0.39-1.32)	.29
4601-8000	0.63 (0.31-1.26)	.19	0.87 (0.48-1.60)	.66
8001-17,000	1.58 (0.76-3.31)	.22	1.25 (0.64-2.45)	.51
>17,000	0.73 (0.28-1.89)	.52	0.66 (0.29-1.50)	.32
Internet (ref: Inactive)	1.10 (0.65-1.86)	.73	0.76 (0.49-1.17)	.22
Trust in health website sponsor (ref: Governmental)				
Organizational	0.79 (0.44-1.44)	.44	0.41 (0.27-0.62)	<.001
Commercial	3.61 (1.87-6.97)	<.001	0.28 (0.15-0.54)	<.001
Personal	0.54 (0.16-1.76)	.30	0.37 (0.16-0.89)	.03
Other	1.51 (0.63-3.61)	.35	0.32 (0.14-0.74)	.01
City (ref: Hangzhou)				
Jinhua	0.53 (0.31-0.91)	.02	0.58 (0.36-0.92)	.02
Lishui	0.88 (0.50-1.55)	.66	0.72 (0.45-1.15)	.17
Hospital type (ref: Primary)				
Secondary B	0.68 (0.20-2.30)	.54	1.64 (0.47-5.66)	.44
Secondary A	0.27 (0.10-0.74)	.01	1.24 (0.43-3.53)	.69
Tertiary B	0.46 (0.19-1.10)	.08	1.69 (0.67-4.26)	.27
Tertiary A	0.33 (0.15-0.73)	.01	1.51 (0.64-3.54)	.34
Constant	3.52 (0.78-15.91)	.10	0.64 (0.15-2.78)	.55

[a] RRR: relative risk ratio.

[b] ref: reference group.

[c] Not applicable.

Discussion

Principal Findings

Chinese patients who were online medical information seekers had a high intention to use virtual visits of internet hospitals. Gender, education, monthly income, and consumer type were significantly associated with the intention to use virtual visits. Patients had different intentions to use virtual visits delivered by different sponsorship types of internet hospitals, in which the public hospital–sponsored one was the most prevalent one, followed by the government, digital health companies, medical e-commerce companies, private hospitals, and other companies. Gender, age, education, city income level, consumer type, and trust in health website sponsor were significantly associated with the patient's intention to use virtual visits delivered by 3 different sponsorship types of internet hospitals (enterprise-sponsored, hospital-sponsored, and government-sponsored).

Comparison With Prior Work

Intention to Use Virtual Visits

This study, to our knowledge, was the first survey that examined the intention to use virtual visits delivered by different sponsorship types of internet hospitals among patients in China. Previous studies have examined the intention to use internet hospitals in China; however, most of them have not specifically examined the intention to use virtual visits of internet hospitals [12,13]. This study revealed that 87.06% of patients who were online medical information seekers had intention to use virtual visits of internet hospitals, which was much higher than the finding by Li et al [12] that showed 65.6% of participants were willing to use internet hospitals. Our statistical analysis examined online medical information seekers' willingness to use virtual visits; therefore, it was very likely that the intention to use was much higher than the study by Li et al [12]. This large difference could be explained partly by the place where the survey has been conducted, since our survey was conducted in Zhejiang Province, while the survey by Li et al [12] was conducted in Sichuan Province, whose internet development index lagged far behind Zhejiang Province [40]. This large difference could also be possibly explained by the measurement of intention to use because we pointed out that the government has issued regulation guidelines for internet hospitals, especially virtual visits, in our question to measure the intention to use virtual visits.

Intention to Use Virtual Visits Delivered by Different Sponsorship Types of Internet Hospitals

Findings of this study have confirmed Hypothesis 2 that patients had different intentions to use virtual visits delivered by different sponsorship types of internet hospitals, as well as Hypothesis 1 that trust in sponsor was the significant determinant of the patient's intention to use virtual visits delivered by different sponsorship types of internet hospitals. Public hospitals and the government were the top 2 sponsors of internet hospitals in which patients had high intention to use virtual visits, which was consistent with previous studies that found patients tended to trust health websites sponsored by hospitals, universities, government agencies, and well-known nonprofit organizations [32,41]. Enterprise-sponsored internet hospitals lagged behind public hospital–sponsored and government-sponsored ones, which was also consistent with prior studies that found that patients were likely to distrust health websites that appeared to be commercial [42-44]. These findings indicated that TAM was the robust model to explain the patient's acceptance of different sponsorship types of internet hospitals.

Three different categories of internet hospitals (enterprise-sponsored, hospital-sponsored, and government-sponsored) varied in target customer, motivation, and online health care resource allocation. The government-sponsored internet hospital aims to deliver internet hospital services to all its local residents by pooling almost all its regional public medical institutions. Hospital-sponsored internet hospitals usually confine the health care resource to the individual hospital itself or the medical alliance brought by the integrated delivery system which often includes public hospitals and primary health care institutions [1]. Enterprise-sponsored internet hospitals aim to deliver internet hospital services to consumers nationwide by attracting licensed physicians nationwide, especially from public tertiary hospitals. As noted, despite the difference in target consumer, motivation, and online health care resource allocation, most internet hospitals recruited most of their physicians from the public health care system. There are definitely competitions between different categories, as well as competitions within categories.

Trust (also often referred to as credibility) has 2 primary components: trustworthiness (perceived motivation) and expertise (perceived ability) to provide accurate and truthful information [45]. Patients had the highest intention to use virtual visits delivered by public hospital–sponsored internet hospitals (52.91%) but very low intention for private hospital–sponsored ones (1.45%). This finding was in accordance with the fact that private hospital–sponsored internet hospitals only accounted for 10% of hospital-sponsored internet hospitals [5]. Because public and private hospitals varied in motivations, patients usually perceived private hospitals as less trustworthy, despite the same expertise level. Furthermore, private hospitals in China lacked the accumulation of reputation, patients had impressions that private hospitals were more concerned about economic benefits rather than patients' benefits, and they fulfilled less social responsibility than public hospitals; tendentious news reports also exacerbated these impressions [46].

Patients had higher intention to use virtual visits delivered by digital health company–sponsored internet hospitals (12.98%) than medical e-commerce company-sponsored ones (5.37%) and other companies (0.67%), which was in accordance with the fact that digital health companies played a dominant place in the enterprise-sponsored internet hospitals (44%) [5]. As noted, digital health companies had particular advantage over all any other companies on the perceived expertise dimension of trust, and thus patients were more likely to trust digital health companies than all other companies, despite the same trustworthiness level. All other companies recently have made efforts to increase their roles in the internet hospital market, especially after the outbreak of COVID-19. Internet tech companies and medical e-commerce companies possessed their critical advantages on the larger user base of their parent firms and easy access to internet end users; we need to wait to see the results of competition within the category of enterprise-sponsored internet hospitals.

Determinants of Intention to Use Virtual Visits Delivered by Different Sponsorship Types of Internet Hospitals

The findings on the association of sociodemographic variables and the intention to use telemedicine remained mixed. This study identified that gender, education, income, and consumer type were the factors significantly associated with the intention to use virtual visits. Most previous studies also found that patients who were more willing to use telemedicine tended to have a higher education level [12,47,48]. The majority of prior studies demonstrated that gender had no significant effect on the intention to use telemedicine [12,48]. This study found that females were more likely to have intention to use virtual visits; one possible explanation was that females were more likely to be the family caregiver [49]. Compared to low-income patients and primary hospitals consumers, high-income patients and tertiary hospitals consumers were more likely to have intention to use virtual visits because they had higher demand for high quality of care and internet hospitals provided them an alternative access to high quality care virtually.

This study indicated that gender, education, and consumer type were the factors significantly associated with both the intention to use virtual visits and the intention to use virtual visits delivered by different sponsorship types of internet hospitals. Patients with the characteristics of being female, higher education level, tertiary A and B hospital, and secondary A hospital consumers were more likely to have intention to use virtual visits delivered by hospital-sponsored internet hospitals other than enterprise-sponsored ones. These groups of patients had higher demand for high-quality care in which hospital-sponsored internet hospitals were perceived as a better deliverer than enterprise-sponsored ones.

Females were more likely than males to prefer the government-sponsored internet hospitals to hospital-sponsored ones, as on one hand the government-sponsored one was more convenient for females (the primary family caregiver) because it pooled the regional public hospitals together; on the other hand, female Chinese were more likely to trust in governments [50]. Patients aged 30 to 40 years were more likely than those aged 18 to 29 years to have intention to use the government-sponsored internet hospital; one possible explanation was that they attached more importance to convenience.

Compared to enterprise-sponsored and government-sponsored internet hospitals, patients in medium-income cities were more likely than those in high-income cities to have intention to use virtual visits delivered by hospital-sponsored ones. Zhejiang Province has done a good job in market penetration, since there are quite a few internet hospitals at municipal level and even at county level [5]. The more developed the city is, more tertiary hospitals the city has. However, virtual visits only cover some common conditions and chronic diseases for non–first-visits; the demand of the patient in a medium-income city would be met by the city's public tertiary hospitals, many of which have set up internet hospitals.

Limitations and Future Research

This study has several limitations. First, this study specifically examined online medical information seekers' intention to use virtual visits delivered by internet hospitals, which might overestimate the whole group of patients' intentions to use. Second, there are urban-rural divides at various aspects in China. The survey was conducted in cities of Zhejiang Province. As rural population was not included in this survey, the findings of this study only revealed urban patients' behavioral intention toward virtual visits. We will extend our study by conducting the survey on the whole population (both urban and rural population, both online and nononline medical information seekers) in the future.

Third, Zhejiang provincial internet hospital platform was initially launched on Alipay (the digital payment platform of Alibaba Group) in January 2019. Our survey was conducted in May and June 2019, and in July 2019, the Zhejiang provincial internet hospital platform was transferred to the all-in-one digital provincial government website/app. This transfer might have some effects on patients' intention to use the government-sponsored internet hospital platform.

Fourth, this study was conducted before the outbreak of COVID-19, which may have been greatly different from the current situation. The outbreak of COVID-19 has brought tremendous growth of internet hospitals in China and has drastically increased the awareness of virtual visits. The rapid increase of awareness would greatly increase the patient's intention to use and actual use of virtual visits. We will follow and further the study by continuous survey on patients' behavioral beliefs, intention to use, and actual use of internet hospitals to measure the impacts of drastic expansion of different sponsorship types of internet hospitals (especially public hospital–sponsored ones) brought by the outbreak of COVID-19.

Conclusion

This study has demonstrated that virtual visits differ from virtual consultations in both service scope and health insurance coverage. The virtual visit market is an emerging market in China. This study has implied that different sponsorship types of internet hospitals (ie, enterprises, hospitals, and the government) have different motivations, target consumers, and online health care resource allocation. This study found that Chinese patients who were online medical information seekers had high intention to use virtual visits. Public hospitals, the government, and digital health enterprises were the top 3 sponsorship types of internet hospitals in which patients had intention to use their virtual visit services.

This study revealed that trust in a health website sponsor significantly influenced the patient's intention to use virtual visits delivered by different types of internet hospitals, which extended the current knowledge regarding the impact of trust on adoption of direct-to-consumer telemedicine service. This study implied that internet hospitals should pay more attention to consumers with characteristics of being female and a tertiary hospital consumer with higher education level and high income level when developing the virtual visit market. It also implied that different sponsorship types of internet hospitals had different target consumers for virtual visit service. Hospital-sponsored internet hospitals had an advantage over the government-sponsored and enterprise-sponsored ones when developing the virtual visit market in medium-income cities.

Acknowledgments

This work was funded by grant 18YJC630100 from the Ministry of Education Project of Humanities and Social Sciences and grant LY19G030007 from the Natural Science Foundation of Zhejiang Province, China.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire (in Chinese).
[DOC File , 42 KB - jmir_v23i8e25978_app1.doc]

References

1. Wu D, Hesketh T, Shu H, Lian W, Tang W, Tian J. Description of an online hospital platform, China. Bull World Health Organ 2019 Aug 01;97(8):578-579 [FREE Full text] [doi: 10.2471/BLT.18.226936] [Medline: 31384077]
2. [Leading Internet Plus healthcare, the hospital won the title of "Pilot Project of Smart Hospital Construction in Guangdong Province"]. 2017 Dec 19. URL: https://www.gd2h.com/news/yydt/a_102757.html [accessed 2020-10-07]
3. [State Council issues guideline on Internet Plus healthcare]. General Office of the State Council of the People's Republic of China. 2018 Apr 28. URL: http://www.gov.cn/zhengce/content/2018-04/28/content_5286645.htm?trs=1 [accessed 2020-10-07]
4. [Three regulation schemes on internet diagnosis and treatment, etc (for trial implementation)]. National Health Commission of the People's Republic of China and the National Administration of Traditional Chinese Medicine. 2018 Jul 17. URL: http://www.gov.cn/gongbao/content/2019/content_5358684.htm [accessed 2020-10-07]
5. Zhang X. [Comprehensive analysis of 479 internet hospitals in China, hospital-sponsored ones accounted for 80% with a declining increase post COVID-19 outbreak]. 2020 May 07. URL: https://vcbeat.top/OTU2NDIwYmFjYmViYWYxZWUzYWRmMWIwYTk5OTNhYmQ= [accessed 2020-10-07]
6. Neufeld JD, Doarn CR, Aly R. State policies influence Medicare telemedicine utilization. Telemed J E Health 2016 Jan;22(1):70-74. [doi: 10.1089/tmj.2015.0044] [Medline: 26218148]
7. [Guidelines to improve price and health insurance reimbursement policies for Internet Plus Healthcare services]. National Healthcare Security Administration. 2019 Aug 30. URL: http://www.nhsa.gov.cn/art/2019/8/30/art_14_1705.html [accessed 2020-10-07]
8. Han Y, Lie RK, Guo R. The internet hospital as a telehealth model in China: systematic search and content analysis. J Med Internet Res 2020 Jul 29;22(7):e17995 [FREE Full text] [doi: 10.2196/17995] [Medline: 32723721]
9. [China’s first internet hospital was built in Wuzhen, and it issued the first online prescription]. 2015 Dec 10. URL: https://36kr.com/p/1720990187521 [accessed 2020-08-08]
10. [Internet hospital of the First Affiliated Hospital of Zhejiang University opens today, the first public hospital sponsored internet hospital in China]. 2016 Feb 17. URL: http://health.people.com.cn/n1/2016/0217/c398004-28131507.html [accessed 2020-10-07]
11. [China's first "service + supervision" integrated Internet hospital platform is officially launched!]. 2019 Jan 22. URL: https://www.sohu.com/a/290809019_120056551 [accessed 2020-10-07]
12. Li P, Luo Y, Yu X, Wen J, Mason E, Li W, et al. Patients' perceptions of barriers and facilitators to the adoption of e-hospitals: cross-sectional study in western China. J Med Internet Res 2020 Jun 11;22(6):e17221 [FREE Full text] [doi: 10.2196/17221] [Medline: 32525483]
13. Zhou M, Zhao L, Kong N, Campy KS, Qu S, Wang S. Factors influencing behavior intentions to telehealth by Chinese elderly: an extended TAM model. Int J Med Inform 2019 Jun;126:118-127. [doi: 10.1016/j.ijmedinf.2019.04.001] [Medline: 31029253]
14. Li D, Hu Y, Pfaff H, Wang L, Deng L, Lu C, et al. Determinants of patients' intention to use the online inquiry services provided by internet hospitals: empirical evidence from China. J Med Internet Res 2020 Oct 29;22(10):e22716 [FREE Full text] [doi: 10.2196/22716] [Medline: 33006941]
15. Shi Z, Mehrotra A, Gidengil CA, Poon SJ, Uscher-Pines L, Ray KN. Quality of care for acute respiratory infections during direct-to-consumer telemedicine visits for adults. Health Aff (Millwood) 2018 Dec;37(12):2014-2023 [FREE Full text] [doi: 10.1377/hlthaff.2018.05091] [Medline: 30633682]

16. Viers BR, Pruthi S, Rivera ME, O'Neil DA, Gardner MR, Jenkins SM, et al. Are patients willing to engage in telemedicine for their care: a survey of preuse perceptions and acceptance of remote video visits in a urological patient population. Urology 2015 Jun;85(6):1233-1239. [doi: 10.1016/j.urology.2014.12.064] [Medline: 25863832]
17. Jain T, Mehrotra A. Comparison of direct-to-consumer telemedicine visits with primary care visits. JAMA Netw Open 2020 Dec 01;3(12):e2028392 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.28392] [Medline: 33289842]
18. Uscher-Pines L, Mulcahy A, Cowling D, Hunter G, Burns R, Mehrotra A. Access and quality of care in direct-to-consumer telemedicine. Telemed J E Health 2016 Apr;22(4):282-287. [doi: 10.1089/tmj.2015.0079] [Medline: 26488151]
19. Tao D, Wang T, Wang T, Zhang T, Zhang X, Qu X. A systematic review and meta-analysis of user acceptance of consumer-oriented health information technologies. Comput Human Behav 2020 Mar;104:106147. [doi: 10.1016/j.chb.2019.09.023]
20. Rahimi B, Nadri H, Lotfnezhad Afshar H, Timpka T. A systematic review of the technology acceptance model in health informatics. Appl Clin Inform 2018 Dec;9(3):604-634 [FREE Full text] [doi: 10.1055/s-0038-1668091] [Medline: 30112741]
21. Harst L, Lantzsch H, Scheibe M. Theories predicting end-user acceptance of telemedicine use: systematic review. J Med Internet Res 2019 May 21;21(5):e13117 [FREE Full text] [doi: 10.2196/13117] [Medline: 31115340]
22. Gefen, Karahanna, Straub. Trust and TAM in online shopping: an integrated model. MIS Quarterly 2003;27(1):51. [doi: 10.2307/30036519]
23. Kamal SA, Shafiq M, Kakria P. Investigating acceptance of telemedicine services through an extended technology acceptance model (TAM). Technol Soc 2020 Feb;60:101212. [doi: 10.1016/j.techsoc.2019.101212]
24. Velsen LV, Tabak M, Hermens H. Measuring patient trust in telemedicine services: development of a survey instrument and its validation for an anticoagulation web-service. Int J Med Inform 2017 Dec;97:52-58. [doi: 10.1016/j.ijmedinf.2016.09.009] [Medline: 27919395]
25. 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]
26. Eysenbach G, Köhler C. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. BMJ 2002 Mar 9;324(7337):573-577 [FREE Full text] [Medline: 11884321]
27. Sun Y, Zhang Y, Gwizdka J, Trace CB. Consumer evaluation of the quality of online health information: systematic literature review of relevant criteria and indicators. J Med Internet Res 2019 May 02;21(5):e12522 [FREE Full text] [doi: 10.2196/12522] [Medline: 31045507]
28. Rice R, Peterson M, Christine R. A comparative features analysis of publicly accessible commercial and government health database web sites. In: Rice RE, Katz JE, editors. The Internet and Health Communication: Expectations and Experiences. Thousand Oaks: Sage Publication; 2001:213-232.
29. West D, Miller E. Digital Medicine: Health Care in the Internet Era. Washington: Brookings Institution Press; 2009.
30. Dutta-Bergman M. Trusted online sources of health information: differences in demographics, health beliefs, and health-information orientation. J Med Internet Res 2003;5(3):e21 [FREE Full text] [doi: 10.2196/jmir.5.3.e21] [Medline: 14517112]
31. Groselj D. A webometric analysis of online health information: sponsorship, platform type and link structures. Online Information Review 2014 Feb 25;38(2):209-231. [doi: 10.1108/OIR-01-2013-0011]
32. Kim Y. Trust in health information websites: a systematic literature review on the antecedents of trust. Health Informatics J 2016 Jun;22(2):355-369. [doi: 10.1177/1460458214559432] [Medline: 25518944]
33. [Zhejiang Statistical Yearbook 2019]. Zhejiang Provincial Bureau of Statistics. URL: https://zjjcmspublic.oss-cn-hangzhou-zwynet-d01-a.internet.cloud.zj.gov.cn/jcms_files/jcms1/web3077/site/flash/tjj/Reports1/2020%E7%BB%9F%E8%AE%A1%E5%B9%B4%E9%89%B420200929/2019%E5%B9%B4%E7%BB%9F%E8%AE%A1%E5%B9%B4%E9%89%B4%E5%85%89%E7%9B%9820200929/indexch.htm [accessed 2020-07-20]
34. [Internet Development Report of Zhejiang Province 2019 was published]. 2020 May 14. URL: http://www.cac.gov.cn/2020-05/14/c_1591019269867594.htm [accessed 2020-08-20]
35. 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. Lancet 2019 Sep 28;394(10204):1192-1204. [doi: 10.1016/S0140-6736(19)32136-1] [Medline: 31571602]
36. Luo M. [China speeds up approval and supervision of internet hospitals since August 2018, over 100 internet hospitals founded, with over 200 firms involved]. 2019 Jun 4. URL: https://vcbeat.top/ZjY1NWM5ZDMzZTc5MmY2ZjJiZWViZTQ0MWFhOTNjYzA= [accessed 2020-11-10]
37. Leung L, Chen C. E-health/m-health adoption and lifestyle improvements: exploring the roles of technology readiness, the expectation-confirmation model, and health-related information activities. Telecomm Policy 2019 Jul;43(6):563-575. [doi: 10.1016/j.telpol.2019.01.005]
38. Rising CJ, Jensen RE, Moser RP, Oh A. Characterizing the US population by patterns of mobile health use for health and behavioral tracking: analysis of the National Cancer Institute's health information national trends survey data. J Med Internet Res 2020 May 14;22(5):e16299 [FREE Full text] [doi: 10.2196/16299] [Medline: 32406865]

39. Wilson EV, Lankton NK. Modeling patients' acceptance of provider-delivered e-health. J Am Med Inform Assoc 2004;11(4):241-248 [FREE Full text] [doi: 10.1197/jamia.M1475] [Medline: 15064290]
40. [The Blue Book reveals the latest development achievements of the Internet. The World Internet Development Report 2019 and the China Internet Development Report 2019 released]. 2020 Jun 12. URL: http://www.cac.gov.cn/2020-06/12/c_1593513279332818.htm [accessed 2020-10-17]
41. Freeman KS, Spyridakis JH. An examination of factors that affect the credibility of online health information. Tech Commun 2004;51(2):239-263.
42. Kerr C, Murray E, Stevenson F, Gore C, Nazareth I. Internet interventions for long-term conditions: patient and caregiver quality criteria. J Med Internet Res 2006;8(3):e13 [FREE Full text] [doi: 10.2196/jmir.8.3.e13] [Medline: 16954123]
43. Fox S, Duggan M. Health online 2013. Washington: Pew Internet and American Life Project; 2013 Jan 15. URL: https://www.pewresearch.org/internet/wp-content/uploads/sites/9/media/Files/Reports/PIP_HealthOnline.pdf [accessed 2021-07-21]
44. Colombo C, Mosconi P, Confalonieri P, Baroni I, Traversa S, Hill SJ, et al. Web search behavior and information needs of people with multiple sclerosis: focus group study and analysis of online postings. Interact J Med Res 2014;3(3):e12 [FREE Full text] [doi: 10.2196/ijmr.3034] [Medline: 25093374]
45. Hovland C, Janis I, Kelley H. Communication and Persuasion: Psychological Studies of Opinion Change. New-Haven: Yale University Press; 1953.
46. Deng C, Li X, Pan J. Private hospital expansion in China: a global perspective. Glob Health J 2018 Jun;2(2):33-46 [FREE Full text] [doi: 10.1016/S2414-6447(19)30138-1] [Medline: 32501412]
47. Turner JW, Thomas RJ, Gailiun M. Consumer response to virtual service organisations: the case of telemedicine. J Med Market 2001 Jul 01;1(4):309-318. [doi: 10.1057/palgrave.jmm.5040042]
48. Scott KC, Karem P, Shifflett K, Vegi L, Ravi K, Brooks M. Evaluating barriers to adopting telemedicine worldwide: a systematic review. J Telemed Telecare 2018 Jan;24(1):4-12 [FREE Full text] [doi: 10.1177/1357633X16674087] [Medline: 29320966]
49. Lilly MB, Laporte A, Coyte PC. Labor market work and home care's unpaid caregivers: a systematic review of labor force participation rates, predictors of labor market withdrawal, and hours of work. Milbank Q 2007 Dec;85(4):641-690 [FREE Full text] [doi: 10.1111/j.1468-0009.2007.00504.x] [Medline: 18070333]
50. Steinhardt HC. How is high trust in China possible? Comparing the origins of generalized trust in three Chinese societies. Political Studies 2011 Nov 18;60(2):434-454. [doi: 10.1111/j.1467-9248.2011.00909.x]

Abbreviations

GDP: gross domestic product
HINTS: Health Information National Trends Survey
mHealth: mobile health
OR: odds ratio
TAM: technology acceptance model

Edited by R Kukafka; submitted 23.11.20; peer-reviewed by S Qu, W Tang, H Hu; comments to author 15.03.21; revised version received 22.04.21; accepted 05.07.21; published 13.08.21.

Please cite as:
Liu L, Shi L
Chinese Patients' Intention to Use Different Types of Internet Hospitals: Cross-sectional Study on Virtual Visits
J Med Internet Res 2021;23(8):e25978
URL: https://www.jmir.org/2021/8/e25978
doi:10.2196/25978
PMID:34397388

©Liyun Liu, Lizheng Shi. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Integrating Welfare Technology in Long-term Care Services: Nationwide Cross-sectional Survey Study

Hanne Marie Rostad[1], PhD; Randi Stokke[1], PhD

Department of Health Sciences in Gjøvik, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Gjøvik, Norway

Corresponding Author:
Hanne Marie Rostad, PhD
Department of Health Sciences in Gjøvik
Faculty of Medicine and Health Sciences
Norwegian University of Science and Technology
Teknologivegen 22
Gjøvik, 2815
Norway
Phone: 47 61 13 54 00
Email: hanne.m.rostad@ntnu.no

Abstract

Background: Welfare technologies are often described as a solution to the increasing pressure on primary health care services. However, despite initiating welfare technology projects in the health care sector and different government incentives, research indicates that it is difficult to integrate welfare technology innovations in a complex and varying setting, such as long-term care.

Objective: We aim to describe the types of welfare technology and the extent to which welfare technology is provided in long-term care (ie, nursing homes and home care services); examine whether the extent of welfare technology provision differs on the basis of municipal characteristics (ie, population size, centrality, the proportion of older inhabitants, and income); and identify how local governments (ie, municipalities) describe their efforts toward integrating welfare technologies in long-term care.

Methods: Quantitative and qualitative data about welfare technology from a larger cross-sectional survey about the provision of long-term care services in Norwegian municipalities were combined with registry data. Representatives of 422 Norwegian municipalities were invited to participate in the survey. Frequencies were used to describe the distribution of the types and extent of welfare technologies, whereas the Fisher exact test and Kruskal-Wallis one-way analysis of variance were used to determine the association between the extent of welfare technology and municipal characteristics. Free-form text data were analyzed using thematic analysis.

Results: A total of 277 municipalities were surveyed. Technology for safety was the most widespread type of welfare technology, whereas technology for social contact was the least prevalent. Two-thirds of the sample (183/277, 66.1%) in nursing home and (197/277, 71.1%) in home care services reported providing one or two different types of welfare technology. There was a statistically significant association between the extent of welfare technology and population size (in both nursing homes and home care services: P=.01), centrality (nursing homes: P=.01; home care services: P<.001), and municipal income (nursing homes: P=.02; home care services: P<.001). The extent of welfare technology was not associated with the proportion of older adults. The municipalities described being in a piloting phase and committing to future investment in welfare technology. Monetary resources were allocated, competency development among staff was initiated, and the municipalities were concerned about establishing collaborations within and between municipalities. Home care services seem to have a more person-centered approach in their efforts toward integrating welfare technologies, whereas nursing homes seem to have a more technology-centered approach.

Conclusions: Many municipalities provide welfare technologies; however, their extent is limited and varies according to municipal characteristics. Municipal practices still seem dominated by piloting, and welfare technologies are not fully integrated into long-term care services. Innovation with welfare technology appears top-down and is influenced by national policy but also reflects creating a *window of opportunity* through the organization of municipal efforts toward integrating welfare technology through, for example, collaborations and committing personnel and financial resources.

(J Med Internet Res 2021;23(8):e22316) doi:10.2196/22316

KEYWORDS

ambient assisted living; cross-sectional survey; home care services; innovation; long-term care; nursing homes; telecare; welfare technology; mobile phone

Introduction

Background

Owing to demographic changes and prolonged living expectations, most Western societies continue to face an increasing population of older adults [1,2]. This has contributed to a policy focus on active and independent living for older and frail adults [3,4]. In addition, the decentralization of health care services and the financial pressure to limit public costs have led to increasing pressure to develop sustainable primary health care services [5-7] because the status quo cannot be maintained [8].

As health care systems and terminology vary across countries, we must clarify the terms *primary health care* and *long-term care*. We define *primary health care* as a broad term that covers health care services throughout an individual's life span, ranging from prevention to treatment, rehabilitation, and palliative care [9]. In Nordic countries, primary health care is predominantly publicly funded at the municipal level [10], the atomic unit of local government in Norway; municipalities vary significantly in size, topography, and demography, resulting in different priorities in the provision of primary health care services.

Long-term care is provided in municipal primary health care and involves services specifically directed at people who need assistance to perform basic activities of daily living. Various services are provided by different caregivers to address both medical and nonmedical needs at home, in assisted living facilities, or in nursing homes [11]. In Norway, municipalities are legally responsible for financing and providing long-term care services, such as home care and nursing homes—services found in every municipality—that are primarily funded through so-called *unrestricted revenues per capita*, consisting of tax revenues and block grants from the central government.

For governments in Western countries, innovation offers a potential *solution* to the aging population, the diminishing workforce, and increased demand for primary health care services [12]. Innovation is often described as a new product or service that represents a significant change for the people involved [13], which is integrated into practice and can be repeated and translated into new contexts [14,15]. Different technological devices are important examples of innovation and are both products and integrated parts of the service through innovation processes. Welfare technologies interact with the people involved in the service; they not only support care, but also change how care is provided and different people's roles and responsibilities [16].

Although many terms can be used to reference technological innovations, in this paper, we refer to them as *welfare technology*. *Welfare technology* is an umbrella term, mainly used in Nordic countries, that covers technologies that have the potential to maintain or improve individuals' functioning, safety, and independence, thereby promoting their well-being and reducing the need for formal and informal care. Other commonly used terms are *telecare*, *telehealth*, *ambient assisted living technologies*, *telemedicine*, and *eHealth*. Although these terms cover different forms of digital care, there is a considerable overlap among them [17,18].

A commitment to welfare technology is prominent throughout the European Union as underpinned by the European Commission's communication on "Telemedicine for the Benefit of Patients, Healthcare Systems and Society" [8]. Western societies and their governments have allocated generous funding to promote technological innovations in care services [17]. This commitment is demonstrated by a series of official government documents in Norway [19-21]. In 2014, the government launched a program for welfare technology innovations in long-term care services—the National Welfare Technology Program—aimed at increasing the focus, investment, and integration of welfare technology in long-term care services [22]. In this paper, the term *integration* is defined as the process of welfare technology becoming a part of the municipality's end-to-end long-term care services to meet the municipality's goals and requirements for long-term care provision. The program is described in detail in Multimedia Appendix 1 [23-25].

Politicians view public innovations as a necessity for prosperity and progress in society [22], and technological innovations in long-term care are high on the agenda of Western societies [26,27]. The Norwegian government aimed to integrate welfare technologies into long-term care services by the end of 2020, and many municipalities have received funding from the government over the last decade to pilot various types of welfare technologies in their care services [28,29]. The current status of welfare technology integration in long-term care services is not known; however, previous research has shown that despite promising results, many welfare technology projects do not pass the pilot stage [30-33]. Consequently, several studies on technology in caring practices have involved pilot studies assessing the drivers of and barriers to integration of such technology into regular use [34,35]. However, these studies do not provide insights into how local governments perceive and organize their efforts toward integrating welfare technology in care services and whether integration varies according to the characteristics of local governments. This study addresses these research gaps using new data from Norwegian municipal long-term care settings. We intend to contribute to the empirical knowledge of welfare technology innovation practices in local governments and provide theoretical insights into how such innovation processes occur in long-term care services.

Theoretical Framing

Studies of technologies emanates from different interdisciplinary academic disciplines [36]. We draw inspiration from Nicolini [37] and on what he describes as a *theory-method package*. This involves zooming in on the practice of integrating welfare technologies in municipal long-term care services and zooming out following trails of connections. Nicolini [37] argues that the purpose of social science is to open a rich and nuanced understanding of the practice and that there is no such thing as a unified practice theory. In this paper, we use what he calls a *toolkit approach* by mobilizing the aspects of different theories when exploring practice, such as theory of public service innovation and those from science and technology studies. Innovation studies and science and technology studies complement each other by focusing on the analysis of the role of technology and the use of technologies in innovation processes [38]. Sharing common origins, both fields have a high degree of interdisciplinarity and focus on how processes unfold, how society is structured, and how innovations evolve in society [39]. A major difference between innovation studies and science and technology studies is that innovation researchers tend to focus and development and look for solutions to problems in management or policy, whereas science and technology studies have a more critical approach, focusing more on the consequences of technology innovation processes [22,40]. Combining these fields enables us to provide an enriched understanding of what is happening. It also provides perspectives and ideas to further explore and understand welfare technology integration in long-term care services in local governments.

Innovations with technology aim to increase the effectiveness and efficiency of the public sector. However, innovation via technology and innovation [41,42] in a long-term care setting and in care work in general [27,43,44] are highly complex, involving *wicked problems* [45,46], such as many and changing stakeholders, competing interests, and disagreements regarding the nature of the problem [47,48]. Star and Ruhleder [16] introduced the importance of infrastructure for technology integration, which was later revisited by Greenhalgh et al [49]. Common metaphors describe infrastructure as “something upon which something else ‘runs’ or ‘operates’” (eg, physical structures, such as railroad tracks upon which rail cars run) [16]. However, Star and Ruhleder [16] and Greenhalgh et al [49] have argued that infrastructure also includes organizational (eg, rules, routines, processes, practices, and norms) and relational (eg, the relationships between the actors involved, as well as the practices and technologies) features, which generate particular agendas and priorities that influence technological innovation and its integration.

Aim and Objectives

This study aims to provide knowledge on the current status of welfare technology integration in long-term care services and how local governments perceive and organize their efforts toward integrating welfare technology.

The study objectives are three-fold: (1) to describe the types and extent to which welfare technology is provided in long-term care; (2) to examine whether the extent of welfare technology provision differs on the basis of municipal characteristics (ie, population size, centrality, the proportion of older inhabitants, and income); and (3) to identify how local governments (ie, municipalities) describe their efforts toward integrating welfare technologies in long-term care.

Methods

In this study, we applied data triangulation, which is the “use of different sources of data as distinct from using different methods in the production of data” [50]. Quantitative and qualitative data from a cross-sectional survey of long-term care settings (ie, nursing homes and home care services) in Norway were combined with registry data from publicly available national statistics.

Setting and Participants

When we conducted our study in 2019, Norway had 422 municipalities distributed among five regions; however, this number declined to 356 municipalities in 2020 due to regional and municipal reforms [51]. Total population sampling was used, a type of purposive sampling technique [52], in which all Norwegian municipalities were invited to participate. The municipalities were contacted via email and asked to choose one person with extensive knowledge of the municipality’s long-term care services to answer the survey on behalf of their municipality. The title of the selected person varied due to the varying organizational structures among municipalities. The varying titles meant varying functions, tasks, responsibilities, and the degree to which the respondent could make decisions in their organization: some made long-term decisions on a strategic level, whereas others decided day-to-day operations on an operational level. The person designated as the respondent was contacted via email, which provided information about the study and participation, in addition to a link to the survey. To encourage participation, the email described how the study could contribute to improved knowledge and understanding of current practices, which could aid both local and national policy makers who plan, set priorities, and develop services for the future, ultimately improving long-term care services for patients and their families. Email reminders were sent three times, approximately every 4 weeks, during the study period.

Data Sources

A web-based questionnaire was designed based on a comprehensive review of the literature. We developed the questionnaire together with a user panel consisting of representatives from five municipalities of different sizes (based on population) and geographical locations. The representatives held different positions as leaders and advisers at different organizational levels but all worked in or for long-term care services. In addition, the representatives in the panel helped establish face validity and reviewed whether the questionnaire effectively captured the topic under investigation and checked for double, confusing, or misleading questions. None of the representatives in the panel participated in the study on behalf of their municipality. In addition to the user panel, the survey was piloted by 3 representatives from the target group, 2 of which also answered the main survey, and their responses were included in the analyses. Adjustments were made to the survey based on the feedback from the user panel and pilot. Specifically, questions were rephrased or removed due to a lack of relevance.

A more detailed description of the questionnaire is provided in Multimedia Appendix 2. As we used conditional branching in the survey, the respondents may take different paths through the survey depending on their answers; not everyone received all the questions. The participants were able to start, stop, and resume answering the questionnaire until the survey was closed. Two close-ended questions were analyzed in this study: what types of welfare technology were provided in (1) nursing homes and (2) home care services? Two open-ended questions allowed the respondents to elaborate on integrating welfare technology in nursing homes and home care services. Data were collected between February and April 2019.

Variables

Questions about the types of welfare technology in nursing homes and home care services were close-ended questions with a predefined list of answers, wherein the respondents could check off all the choices that applied to them. The predefined list was based on the categorization in the official government document "Innovation in the Care Services" [53] (Textbox 1).

Textbox 1. Categories of welfare technologies.

Categories and examples

- Localization technologies
  - GPS
- Compensation technologies
  - Remote control of light and heating, robot vacuums, and cognitive or physical aids
- Safety technologies
  - Social alarm and fall detection sensors
- Technologies for social contact
  - Tablets, smartphones, gaming, and therapeutic robots
- Treatment technologies
  - Medical remote monitoring and automated pill dispensers

The extent of welfare technology was measured as the number (ranging from 0 to 5) of the different types of welfare technologies provided by the municipality. Data on municipal characteristics included population size, centrality, the proportion of older adults, and municipal income, which were obtained from publicly available statistics (Statistics Norway):

- Population size has three categories, as follows: small (<4999 inhabitants), medium (5000-19,999 inhabitants), and large (>20,000 inhabitants) [54]. Data were from the first quarter of 2019.
- Centrality has three categories, as follows: least central, central, and most central. Data were from January 2018 and based on Statistics Norway's centrality index. The centrality index was based on the travel time to workplaces and service functions (eg, post office and bank) [52].
- The proportion of older adults was a continuous variable for the percentage of the municipality's inhabitants aged ≥80 years in 2019.
- Municipal income was measured as "unrestricted revenues per capita," which is a continuous variable for how much income the municipalities have at their disposal after covering the fixed costs, indicating the municipalities' financial leeway. Data were from the first quarter of 2019.

Data Analysis

Quantitative data were analyzed using SPSS version 25 (IBM Corporation). Frequencies were used to describe the distribution of the types and the extent of welfare technologies provided in nursing homes and home care services. The differences between the responders and nonresponders in terms of municipal characteristics were tested using a chi-square analysis and two-tailed *t* test for independent samples. The Fisher exact test was used to determine the association between the extent of welfare technology and population size and centrality, whereas the Kruskal-Wallis one-way analysis of variance was used to determine the association between the extent of welfare technology and the proportion of older adults and municipal income. The free-form text data from the open-ended survey questions were analyzed using the thematic analysis by Braun and Clarke [53], as described in Multimedia Appendix 3 [53]. Qualitative data were analyzed manually, resulting in the following four themes: (1) from good intentions to established reality, (2) investments in and rigging up the welfare technology initiative, (3) type of technology they are going for, and (4) rationale for focus and selection.

Ethical Considerations

Before the initiation of data collection, the Data Protection Authority within the Norwegian Centre for Research Data assessed the study procedure and concluded that the processing of personal data in this study was in accordance with privacy legislation (reference no. 847216).

An informed consent form was attached to the email sent to the potential respondents, stating that the person consented to participate in the study by completing the survey. Participation was confidential. The participating municipality (not the individual completing the questionnaire) was identified only by the researchers.

Results

Overview

A total of 65.6% (277/422) of municipalities completed the survey. There was a large geographical spread among the responding municipalities, and all five regions were represented (Table 1).

Responders and nonresponders were compared in terms of population size, centrality, proportion of older adults, and municipal income. The smallest municipalities (in terms of population size) were underrepresented (Table 2). The responders had a lower proportion of older adults and lower mean municipal income than nonresponders. In terms of the open-ended questions, there were 170 responses out of 554 possible responses (277 municipalities×2 questions).

Table 1. Geographical spread of responders.

Region	Municipalities in the regions, n (%)	Municipalities participating in the study, n (%)
Northern	87 (20.6)	44 (15.9)
Mid	48 (11.4)	33 (11.9)
Western	120 (28.4)	81 (29.2)
Southern	30 (7.1)	18 (6.5)
Eastern	137 (32.5)	101 (36.5)

Table 2. Comparison of responders and nonresponders (N=422).

Variable	Responders (n=277)	Nonresponders (n=145)	*P* value
Population size, n (%)			.03
Small	131 (47.3)	88 (60.7)	
Medium	105 (37.9)	39 (26.9)	
Large	41 (14.8)	18 (12.4)	
Centrality, n (%)			.20
Most central	23 (8.3)	7 (4.8)	
Central	106 (38.3)	49 (33.8)	
Least central	148 (53.4)	89 (61.4)	
Older inhabitants (%), mean (SD)	5.2 (1.4)	5.6 (1.5)	.01
Municipal income (NOK[a]), mean (SD)	62,567.9 (12,705.6)[b]	66,730.0 (13,956.1)[c]	.01

[a] NOK: Norwegian Kroner.

[b] Equivalent to a mean of US $7669.1 and an SD of US $1557.4.

[c] Equivalent to a mean of US $8179.4 and an SD of US $1710.7.

Types and Extent of Welfare Technology

Independent of setting, technology for safety was the most widespread welfare technology provided, followed by localization technology. Almost all respondents, 96% (266/277), reported having technologies for safety in home care services, whereas 81.9% (227/277) reported them for nursing homes. Examples of technologies for safety include social alarms, fall detectors, bed and chair sensors, and digital supervision. Technology for social contact was the least prevalent, provided by 11.6% (32/277) in nursing homes and 9.7% (27/277) in home care services. Examples of social technology include videoconferences and therapeutic robots.

The thematic analysis of the participants' open-text responses revealed that in the home care setting, the municipalities chose both *types* of technology and selected certain groups to receive the technology:

> *The municipality offers follow-up through telemedicine for patients with chronic obstructive pulmonary disease and diabetes.* [Respondent from a small, least central municipality with an average proportion of older adults aged ≥80 years and a high income]

This quote illustrates how the municipalities directed their focus to particular user groups, such as people with different kinds of chronic diseases.

With regard to the extent of welfare technology, most respondents reported having only one or two different types of welfare technology, irrespective of the setting (Table 3).

Table 3. Types and extent of welfare technologies available.

Variable	Nursing home	Home care services
Types, n (%)		
Localization	137 (49.5)	129 (46.6)
Compensation	48 (17.3)	50 (18.1)
Safety	227 (82)	266 (96)
Social contact	32 (11.6)	27 (9.8)
Treatment	35 (12.6)	73 (26.4)
Extent, n (%)		
None	36 (13)	8 (2.9)
1	88 (31.8)	102 (36.8)
2	95 (34.3)	95 (34.3)
3	40 (14.4)	44 (15.8)
4	13 (4.7)	21 (7.6)
5	5 (1.8)	7 (2.5)

Variations According to Municipal Characteristics

The extent of welfare technology varies according to municipal characteristics, including size, centrality, and income. Surprisingly, our data did not show an association with the proportion of older adults. These variations can be both natural and justifiable, but also contribute to some challenges—something we will come back to in the discussion chapter.

There was a statistically significant association between the extent of welfare technology, population size, and centrality in both nursing homes and home care services. A larger percentage of the largest, most central municipalities seem to provide more types of welfare technologies than the smallest, least central municipalities (Tables 4 and 5).

Table 4. Association between the extent of welfare technology in nursing homes and population size and centrality.

Variable	None	1	2	3	4	5	*P* value
Population size, n (%)							.01
Small	21 (16)	45 (34.4)	45 (34.4)	17 (12.9)	3 (2.3)	0 (0)	
Medium	12 (11.4)	37 (35.2)	37 (35.2)	11 (10.5)	5 (4.8)	3 (2.9)	
Large	3 (7.3)	6 (14.6)	13 (31.7)	12 (29.3)	5 (12.2)	2 (4.9)	
Centrality, n (%)							.01
Most central	3 (13)	2 (8.7)	7 (30.4)	8 (34.8)	3 (13)	0 (0)	
Central	11 (10.4)	34 (32.1)	37 (34.9)	13 (12.3)	6 (5.7)	5 (4.7)	
Least central	22 (14.9)	52 (35.1)	51 (34.5)	19 (12.8)	4 (2.7)	0 (0)	

Table 5. Association between the extent of welfare technology in home care services and population size and centrality.

Variable	None	1	2	3	4	5	*P* value
Population size, n (%)							.01
Small	5 (3.8)	63 (48.1)	4 (3.1)	17 (13)	4 (3.1)	2 (1.5)	
Medium	2 (1.9)	32 (30.5)	41 (39.1)	16 (15.2)	10 (9.5)	4 (3.8)	
Large	0 (0)	8 (19.5)	14 (34.2)	11 (26.8)	7 (17.1)	1 (2.4)	
Centrality, n (%)							<.001
Most central	1 (4.4)	5 (21.7)	4 (17.4)	7 (30.4)	6 (26.1)	0 (0)	
Central	0 (0)	33 (31.1)	43 (40.6)	17 (16)	8 (7.6)	5 (4.7)	
Least central	6 (4.1)	65 (43.9)	48 (32.4)	20 (13.5)	7 (4.7)	2 (1.4)	

There was a statistically significant association between the extent of welfare technology and municipal income in both nursing homes (*P*=.02) and home care settings (*P*<.001). Municipalities with a lower income provided more types of welfare technologies than those with a higher income. No statistically significant association was found between the extent of welfare technology and the proportion of older inhabitants in either nursing homes (*P*=.33) or home care settings (*P*=.50).

From Good Intentions to Established Reality

This theme illustrates how far the municipalities have come in the process of integrating welfare technologies into their long-term care services, which varied considerably. A few municipalities described that welfare technologies were integrated into the municipality's long-term care services, whereas status in most municipalities seemed well reflected in the following quote:

> *We have not integrated these types of solutions, but we take part in the National Welfare Technology Program and will pilot different types of welfare technology in 2019–2020.* [Respondent from a small, least central municipality with an average proportion of older adults and a high income]

Some municipalities described projects focused on updating established technologies, whereas most municipalities participated in projects concerning the integration of new technologies. Few respondents reported that welfare technology was integrated into their care services.

Investing in and Rigging up the Welfare Technology Initiative

This theme concerns how municipalities, as local governments, perceived and organized the integration of welfare technology in their long-term care services. The municipalities appeared enthusiastic about integrating welfare technology as respondents described how they are committing themselves to further initiatives toward integrating more technology, both in terms of extending them to new patient groups and use in other services:

> *Automated pill dispensers will be offered to suitable users in 2019, as well as increased diffusion of safety and sensor technology to more users, for example, for localization, falls, etc...The municipality is now tendering for welfare technology solutions—initially for purchasing new social alarms. More welfare technological solutions will eventually come.* [Respondent from a small, least central municipality with an above average proportion of older adults and a high income]

Other municipalities described how they are preparing for future investments in welfare technology:

> *We have allocated funds in the financial planning period to invest in welfare technology; NOK 500,000 [approximately US $60,000] in 2019, and NOK 1 million [approximately US $120,000] for each of the three upcoming years.* [Respondent from a medium, least central municipality with an average proportion of older adults and a high income]

Some municipalities described how they invest in competency development among staff and that they have staff that are dedicated to working with welfare technology investment regarding mapping needs and what is available:

> *The municipality is at the starting point with welfare technology, with continuing education and basic training of staff.... We have employed a dedicated resource person/adviser in welfare technology.* [Respondent from a large, central municipality with an average proportion of older adults and a low income]

In terms of timing, the investment was done when certain opportunities presented themselves, often involving various demands concerning the necessary attributes of welfare technologies:

> *The municipality is in a phase of constructing new assisted living facilities and a new nursing home, general practitioners' office, physio/occupational therapy, health, and mental health center. A lot of time and resources are used to create good welfare technology solutions for the future. Two people from IT are included in the project group.* [Respondent from a small, least central municipality with an above average proportion of older adults aged ≥80 years and a low income]

> *The acquired welfare technology must be cloud-based, provident, and scalable, and all notifications are received on the same response unit. New sensors/technology will be connected in line with patient needs.* [Respondent from a small, least central municipality with an above average proportion of older adults aged ≥80 years and a low income]

As illustrated above, the municipalities were also concerned with intersectional collaboration, for example, between the municipality's care and information technology sectors. In addition, many municipalities described an intermunicipal collaboration concerning the initiation and execution of welfare technology projects, and some collaborated on assessing the needs for welfare technology investment, inviting tenders, and investing in welfare technology for common use.

Rationale for Investing in Welfare Technology Solutions

The rationale for investing in welfare technologies in nursing homes was different from that for home care services. In home care, the choice of technology depended on an individual patient's needs, whereas in nursing homes, the choice was directed at a collective service *offered* to every resident. This indicates a difference in perspectives and approaches—user-centered versus technology-centered. Home care services were also concerned with who was responsible for what types of technology:

> *We are concerned with the individual's freedom and responsibility to be able to buy technology they feel they need for themselves. We believe that it is not a*

> *matter of course that all welfare technology should be offered by the municipality. Safety technology and technology that can replace supervision (digital supervision)—yes, but automated lighting and heating? No, I mean private builders and private homeowners must be able to take responsibility for themselves.* [Respondent from a medium, central municipality with a below average proportion of older adults and a high income]

In contrast, in nursing homes, technology was linked to the institution, with a focus on general use:

> *The building will be equipped with alert technology, video communication, digital surveillance, and remote monitoring.* [Respondent from a medium, central municipality with an average proportion of older adults and a medium income]

Discussion

Principal Findings

By describing the types and extent to which welfare technology is provided in nursing homes and home care services and determining whether the extent of welfare technology provision differs according to municipal characteristics (the *what* and the *who*), this study aimed to provide knowledge of the current status of welfare technology integration in long-term care services and how local governments perceive and organize their effort toward integrating welfare technology. Furthermore, this study identified how local governments (ie, municipalities) describe their efforts toward integrating welfare technologies in long-term care (the *how* and the *why*).

The What and Who of Technology Innovations in Long-term Care Services

Our findings suggest that welfare technologies are widespread in Norwegian home care services and nursing homes as most municipalities provided them. However, the extent was not great; most municipalities provided only one or two different types of welfare technologies.

The reported welfare technologies for safety provided by the broad number of municipalities most likely refers to social alarms, which are standard in Norway [55] and widely used in other Western societies [56]. Norwegian health authorities recommend municipalities to invest in technologies for safety, along with localization technology, because “the reward is evident” [55]. Accordingly, our findings showed how a national policy initiates local innovation and how a national policy becomes a local priority. Considerable amounts of innovation in public services and frontline work processes have been initiated by the central levels of government with the intention that they are implemented in a top-down manner and disseminated at the frontline level of public service organizations [57]. In this way, national governments can initiate innovation processes through policy documents, regulations, and funding, and the local government, which is closer to citizens, can initiate, develop, and activate the innovation processes [58]. Thus, local governments are both tools for state governance and an arena for local governance [59]. The Norwegian Welfare Technology Program is an example of how the central government uses *soft measures*, such as financial initiatives, to allow local governments to decide whether they will respond to financial stimulants and political signals. Previous research on government reform efforts to establish primary health care services (ie, local medical centers) has indicated that government funding has great significance for local investment and that municipal investments are more influenced by the possibility of state funding than a real need for services [60]. Thus, the question of how central government funding schemes impact real local self-government is relevant for further research.

Our finding that safety technology is the most prevalent type can also be explained by the fact that ensuring patient safety is a top priority in global health [54] and a key expectation for welfare technology from the perspectives of both the government [19] and care professionals [61]. Furthermore, the concept behind welfare technology is remote patient monitoring, focusing on the personal safety of the individual [62]. Although providing welfare technologies is not required by the law, the Norwegian government strongly encourages municipalities to integrate them into their health care services. The services required by law are the municipalities’ main priority, making innovation and service development through welfare technology somewhat secondary. Thus, municipalities would acquire certain types of technologies with clear potential for more efficient services and cost savings, such as safety technologies, rather than social technologies. For instance, digital supervision, such as through fall sensors, can reduce the need for regular checkups in the patient’s room or home and thus the number of staff per shift, whereas a video communication device might not have the same efficiency potential. Furthermore, a study on Norwegian nursing homes found that registered nurses rarely had any time to address the residents’ psychosocial needs because they felt that they had to prioritize their medical and physiological needs [63]. Social needs, such as social contact and belongingness, are basic human needs and are a particular focus of nursing practice and a key task for long-term care services. However, social needs may largely be taken care of by long-term care recipients themselves through devices that are widespread in today’s society, such as smartphones and tablets. As such, the technology for social contact may have already been acquired by the individual care recipient, which may also explain why only a few municipalities provided this category of technology in long-term care.

Following this notion, our findings bring forward an interesting question of *responsibility*, indicating the ongoing debate about the extent to which local governments should be responsible for providing technologies for their inhabitants and what the individuals themselves should acquire. In Nordic countries with large public sectors, there has been little tradition of individual investments in welfare technologies. However, having the current and future users of primary health care services take more responsibility for their own health and well-being is a key goal of an ongoing social reform in Norway called “A Full Life—All Your Life—A Quality Reform for Older Persons” [64], one aim of which is to make the user more accountable by, for example, customizing and equipping their own homes to facilitate independence.

Another finding is the statistically significant association between the extent of welfare technology and population size and centrality in both nursing homes and home care services. The largest and most central municipalities provided more types of different technologies than the smallest and least central municipalities. One possible explanation may be that larger, more central municipalities have a larger number and a more continuous flow of patients with complex chronic diseases, which demand more resources. Thus, providing certain services, such as welfare technology, is likely to be more sustainable in larger, more central municipalities than in those with fewer patients. In addition, larger municipalities have different competence compositions and other resources [65], possibly impacting their capacity to apply for government project funding. Urban areas also have some distinguishing features that are important for innovation capacity: density and diversity [66], where "...density creates a constant flow of new information that comes from observing others successes and failures" [67] and where population diversity means that there are different challenges, needs, resources, experiences, cultural and religious backgrounds, creating an environment wherein *out-of-the-box* ideas are more likely to occur [66]. On the other hand, the benefits of welfare technology may be better appreciated by caregivers operating in rural areas because this technology can remedy large travel distances [68] and provide rural residents better access to chronic disease prevention and quality of care [69,70].

However, the potential benefits of different types of welfare technology differ according to municipal characteristics, such as centrality, population size, and demographics, and the population's health and care needs. For example, the technology for optimizing route planning in home care services is likely more beneficial in large, central municipalities with more traffic and several route options than in small rural municipalities. On the other hand, medical remote monitoring may have more benefits in small rural municipalities. When integrating welfare technologies into different contexts, we expect the same processes and effects. This is problematic because we simplify the complexity and wicked problems. The examples above illustrate how we must always consider the context because welfare technology integration always involves interactions between technology and humans within a context [34].

We also found that municipalities with lower income provided more types of welfare technologies than those with higher incomes. This is likely explained by the fact that innovation in services as an increased use of welfare technology is driven by the strained economy of local governments [71]. Local governments with large financial leeways may not need new or different ways of producing long-term care services because they can afford to maintain the status quo, whereas those with a more constrained financial situation need new solutions to sustain the required standards of care. As such, necessity is the mother of invention.

There is a possible additional aspect concerning the association between the extent of welfare technology and municipal income: can welfare technology be viewed as a form of *second-class* care? In debates regarding welfare technology, some researchers have asked whether *cold* technologies will be integrated at the cost of *warm* human care [72]. Welfare technologies affect the service provided and might be seen as an antagonist to traditional *warm* care. Furthermore, questions have been raised about whether the use of welfare technology creates a stigma because it may signal an inability to master everyday things [73], hence looking and feeling old or vulnerable [74,75], even though the aim is independence. Thus, there may be some reluctance from local stakeholders, such as politicians and leaders of long-term care services, to introduce value-changing care innovations if they are not *forced* to do so by, for example, a strained municipal economy.

The fact that a higher proportion of adults aged above ≥80 years was not associated with the provided welfare technology was somewhat surprising, as welfare technology is proposed as a potential *solution* to the age wave [12] and that much of the welfare technology developed is for this user group. However, this may indicate that *age* is not important for the municipalities' provision of welfare technology. Municipalities that invested in technology aimed at specific user groups, such as patients with dementia and chronic obstructive pulmonary disease, did likely do so due to the individual's physical and cognitive function, rather than the individual's age. Reduced physical and cognitive function results in reduced independence, entailing high costs for the individual, family, and society, whereas health care spending on healthy and independent older adults is relatively modest [76]. Caring for people affected by chronic obstructive pulmonary disease, stroke, and dementia—prevalent patient groups in long-term care—are resource-draining and welfare technologies with potential for more efficient services and cost savings are therefore worth investing in.

As discussed, the extent of welfare technology seems to vary according to municipal characteristics, which is justifiable for several reasons. As the municipalities significantly varied in terms of size, topography, and population composition, different needs must be met when providing care to their inhabitants. Thus, the potential of welfare technology in providing care will depend on many different factors, a limited number of which were explored in this study. However, an important question might be whether the differences in welfare technology provision can threaten the founding principles of many health care systems, such as universality and horizontal equity (ie, equal treatment of individuals or groups in the same circumstances), leading to variations in the quality of care.

The How and Why—the Window of Opportunity

Our results indicate that the municipalities were far from achieving the government's goal of integrating welfare technology into Norwegian long-term care by the end of 2020. Most municipalities described how they were involved in projects or pilot testing. Over the past two decades, there has been a steady growth in the number of technological innovation projects in the health care sector [30]. Both policy makers and researchers have raised concerns over pilots in eHealth and telemedicine, calling for the large-scale integration of technology in routine health service delivery [30]. A *plague of pilots* has been conducted [77], in which projects were established to be run as nonpermanent tests rather than integrated into routine service [30].

One reason for this is that it is always difficult to integrate new technologies within existing organizations, as they complicate complex daily care practices. The integration of new technology implies a change in the interactions and relationships in the organization that brings in new ideas, actors, tasks, and organizational changes in the service. Care work already largely involves actors in the network needing to adjust their practices or technologies [43,74]. In addition, although the municipalities receive incentives from the national government, they have many reasons to be careful when integrating new welfare technologies into their services. There is always a risk of failure that results in losing money with innovations, and municipalities are governed by the rules of democracy and laws [46]. Innovation might also not lead to improvement [14], and although they could lead to improvement, we must always ask "for whom is this an improvement?" In addition, public service has multiple aims that are sometimes conflicting, aiming to provide public and individual values [75].

Although welfare technology does not currently seem integrated into Norwegian long-term services, we interpret the municipalities as being in a *window of opportunity*, which represents a situation in which an established regime becomes unstable (because of external factors or internal problems) and is receptive to alternative regimes and innovation [78]. One example of how municipalities may be viewed as being in a window of opportunity is the current societal and health care reforms and associated financial schemes that encourage municipalities to innovate [79,80], for example, within welfare technology. Other examples are how municipalities are rigging up their welfare technology initiatives: they are entering into intermunicipal collaborations. The collaborative approach in welfare technology initiatives is a major facilitator of possible innovations [81,82]. In addition, collaboration entails shared costs and responsibilities that may result in more efficient and sustainable service solutions, especially for small municipalities [83]. The mode of operation within and among municipalities becomes more comprehensive and differentiated, but also more complex because power may become more elusive as decisions are made in larger networks, decisions for each individual municipality are made through negotiations across several municipalities, and political accountability may be difficult to determine [83]. The health care system is widely recognized as a complex system [84] loaded with wicked problems; thus, innovation in this setting is inherently complicated. The complexity within the system will shape its ability to create a window of opportunity to actively pursue change, adapt, and integrate innovation [85].

Furthermore, this study found that the municipalities are building competence in welfare technology by training and dedicating staff to the welfare technology initiative, indicating an ambition beyond just piloting. This indicates that the municipalities acknowledge that success with technological innovations takes more than just investing in technological devices. Many municipalities are also building and upgrading nursing homes, and they describe these processes as opportunities to invest in welfare technology in long-term care.

Geels and Schot [86] illustrated that transitions and system changes emerge through the interactions among processes at different levels (see the illustration by Geels and Schot [86] in Multimedia Appendix 4 [87]). Considering their theorization, we interpreted our data similarly: how public and technological innovation is shaped by different processes at different societal levels, for example, how demographic changes put pressure on current long-term care practice, thus opening a window of opportunity for innovation, such as welfare technology. This affects the development of national policy and may initiate changes in current practices, such as how the national welfare technology program (as national policy) is taken up by smaller units of actors (ie, municipalities) and further evolved there. The municipalities establish internal momentum through their local learning processes, which are conveyed back to the central government and can lead to adjustments, breakthroughs, and new windows of opportunity. This perspective can contribute to an understanding of how changes and ideas at one level trigger ideas and changes at other levels. This describes how innovation with welfare technology and its integration to its full potential are affected by human agency, social structures, and multilevel interactions.

Another interesting finding adding to the complexity of welfare technology innovations in long-term care is that the integration of welfare technology is characterized by different rationales in nursing homes and home care services. In home care settings, the inclusion of welfare technology was described as dependent on the wishes of the patients, whereas in nursing homes, this person-centeredness focus did not appear, and the focus was more directed at a collective service that every resident was *offered* and at a need for universal design. This can be explained by the different conditions and considerations when integrating welfare technology in nursing homes compared with home care services. Norwegian nursing homes function as a medical and health care institution, and a person moves there for a limited time until they die; in Norway, the median living time in nursing homes is 1.3 years [88], with a steady substitution of patients. Thus, it seems more sustainable to bring in technology linked to the institution rather than the individual patient, as technology in the nursing home setting needs to address universal needs and devices that would *fit* most people living in nursing homes. Furthermore, those who live in nursing homes inevitably face a significant reduction in the range of options available to them by the nature of institutionalized living itself, and an emphasis on the physical care of nursing home residents and a task-oriented approach may result in privileging efficiency over resident choice [89]. Although there has been a fundamental shift in thinking about nursing homes over the past two to three decades—now viewed as person-centered homes offering long-term care services rather than institutions [90], entering a nursing home still entails receiving care in a professional domain compared with home care services, where professionals enter the individuals' private domain to provide care.

In summary, our discussion aims to illustrate how the integration of welfare technology innovations is formed by different factors, processes, and practices at different government levels (ie, national and local) as described by Geels and Schot [86]. These factors, processes, and practices are part of the infrastructure introduced by Star and Ruhleder [16] and Greenhalgh et al [49]. In this study, we analyzed factors such as municipal size and centrality, population factors, and municipal income. Furthermore, we discussed other factors such as policy, national government initiatives, and institutional norms and values, which are also factors included in the infrastructure that may affect the integration of welfare technologies. Focusing on infrastructure provides knowledge of how different factors, processes, and practices are inevitable parts of innovation. This aligns with what we know from the service innovation literature that innovation rarely can be diffused but must be translated to fit the new infrastructure. Thus, the results of studies like ours, focusing on welfare technology integration, will depend on the infrastructure and apply to the particular setting that the study aims to explore.

Limitations and Recommendations for Further Studies

The organization of long-term care differs across countries, thus limiting the generalizability of our findings. However, we believe that our study provides new knowledge and relevant questions when addressing technological innovations in the public care sector in general.

In this paper, we studied the efforts to integrate welfare technology from the perspective of municipal managerial employees based on their knowledge and perceptions of the provision of welfare technology in long-term care services. As the respondents were not engaged in hands-on work, their knowledge of welfare technology provision may be incomplete.

As illustrated by Geels and Schot [86], innovation initiatives also occur in a bottom-up fashion. Therefore, it would be interesting for future studies to elicit the point of view of those who integrate such technologies into the daily operations of long-term care institutions and their users. In addition, this study was neither designed to investigate causality nor study behaviors, opinions, themes, and motivation in-depth. Therefore, in addition to investigating other settings and perspectives, future studies applying other designs and methods will be valuable in explaining and trying to understand the integration of welfare technology into health care services.

Although our sample covered 65.6% (277/422) of the entire Norwegian population of municipalities, small municipalities seemed somewhat underrepresented, whereas those with lower proportion of older adults and lower mean income appeared slightly overrepresented. Furthermore, the data for the open-ended questions were limited (only one-third of what could be obtained), so our results should be used with caution.

The categorization of welfare technology in this paper and Norwegian official government documents entails a significant overlap among the different types of technologies (eg, an automated pill dispenser can be both a treatment technology and a safety technology). Furthermore, the categories defined follow administrative logic and policy issues. Logics represent “frames of reference that condition actors’ choices for sense-making, the vocabulary they use to motivate actions, and their sense of self and identity” [91]. A politician, a long-term care nurse, and a long-term care recipient will all have varying assumptions, values, beliefs, and rules that shape their logic. Thus, the categorization of welfare technology used in this paper may need to be challenged and problematized in future studies because this understanding may have difficulty reaching the core functioning of long-term care when based on a logic emanating from policy goals and administrative rationale.

Considering the ongoing COVID-19 pandemic, studies focusing on whether (and how) the pandemic has affected investments in and use of welfare technologies in different health care settings would be of great interest, for example, whether the use of welfare technologies was introduced (eg, medical remote monitoring and social technologies) or technologies were used in new ways to compensate for the reduced possibility of physical presence in the service recipients’ place of residence, both on the part of formal and informal caregivers.

Conclusions

This study has provided knowledge on the current status of welfare technology integration in long-term care services and insights into how local governments perceive and organize their efforts toward integrating welfare technology in the long-term care setting, thus contributing ideas into how technology innovation processes play out.

Many municipalities provide welfare technologies, whereas most provide safety technologies. However, their extent is limited and varies according to municipal characteristics. Welfare technologies do not seem to be fully integrated into Norwegian long-term care services because many are still in their project and piloting phases. However, the municipalities are motivated to invest in the welfare technology initiative. Therefore, we suggest the term *window of opportunity* as a way of providing insights into the potential for a real integration of welfare technologies rather than solely focusing on *pilotism* as a concern. *Window of opportunity* may be more appropriate for both technology innovations and long-term care services that are complex and innovation processes that take time and opportunity. In addition, innovations are shaped at different societal levels, affecting the ability to create a window of opportunity and facilitate innovation processes.

Acknowledgments

This study is part of the PriCare project funded by the Research Council of Norway (grant 200446).

Authors' Contributions

The first author (HMR; together with a larger research team) planned the study and carried out data collection and the analysis of the close-ended survey questions. Two authors (HMR and RS) analyzed the free-text data. Both authors (HMR and RS) discussed the results, and the first authors drafted the manuscript with contributions from the second author.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The National Welfare Technology Program in Norway.
[DOC File , 41 KB - jmir_v23i8e22316_app1.doc]

Multimedia Appendix 2
Questionnaire.
[DOC File , 45 KB - jmir_v23i8e22316_app2.doc]

Multimedia Appendix 3
The phases of the thematical analysis.
[DOC File , 86 KB - jmir_v23i8e22316_app3.doc]

Multimedia Appendix 4
Geels and Schot processual framework.
[DOC File , 154 KB - jmir_v23i8e22316_app4.doc]

References

1. Tarricone R, Tsouros A. Home Care in Europe. The Solid Facts. WHO Regional Office for Europe. 2008. URL: https://www.euro.who.int/__data/assets/pdf_file/0005/96467/E91884.pdf [accessed 2021-07-01]
2. World Report on Ageing and Health. World Health Organization. 2015. URL: https://apps.who.int/iris/bitstream/handle/10665/186463/9789240694811_eng.pdf?sequence=1&isAllowed=y [accessed 2021-07-01]
3. Boudiny K. 'Active ageing': from empty rhetoric to effective policy tool. Ageing Soc 2013;33(6):1077-1098 [FREE Full text] [doi: 10.1017/S0144686X1200030X] [Medline: 23913994]
4. Foster L, Walker A. Gender and active ageing in Europe. Eur J Ageing 2013;10(1):3-10 [FREE Full text] [doi: 10.1007/s10433-013-0261-0] [Medline: 28804278]
5. Mørk E, Beyrer S, Haugstveit F, Sundby B, Karlsen H. Kommunale Helse- Og Omsorgstjenester 2017. Statistikk Om Tjenester Og Tjenestemottakere [Nursing and Care Services 2017. Statistics on Services and Recipients]. Statistics Norway. 2017. URL: https://www.ssb.no/helse/artikler-og-publikasjoner/_attachment/358290?_ts=165a44eac40 [accessed 2021-07-01]
6. Fuglsang L, Rønning R. Spredning av innovasjon i kommunene [Dissemination of innovation in municipalities]. In: Ringholm T, Teigen H, Aarsæther N, editors. Innovative kommuner [Innovative municipalities]. Oslo: Cappelen Damm Akademisk; 2013:87-102.
7. Vetvik E, Disch PG. Kommunal omsorg - i plan? Status 2011 - utfordringer mot 2025. [Municipal care - in plan? Status 2011 - challgenges towards 2025]. NTNU Open. 2014. URL: https://ntnuopen.ntnu.no/ntnu-xmlui/bitstream/handle/11250/197116/Rapport6_14_web.pdf?sequence=1&isAllowed=y [accessed 2021-07-01]
8. Publications Office of the EU. A new drive for primary care in Europe: Rethinking the assessment tools and methodologies : report of the expert group on health systems performance assessment. Publications Office of the EU. 2018. URL: https://ec.europa.eu/health/sites/health/files/systems_performance_assessment/docs/2018_primarycare_eg_en.pdf [accessed 2021-07-01]
9. World Health Organization. Primary health care. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/primary-health-care [accessed 2021-07-01]
10. Krasnik A, Paulsen B. Recent Reforms and Current Policy Challenges. In: Magnussen J, Vrangbæk K, Saltman RB, editors. Nordic health care systems Recent Reforms and Current Policy Challenges. Milton Keynes: Open University Press; 2009.
11. National Institute of Aging. What Is Long-Term Care? National Institute of Aging. URL: https://www.nia.nih.gov/health/what-long-term-care [accessed 2021-07-01]
12. Huber M, Rodrigues R, Hoffmann F, Gasior K, Marin B. Facts and figures on long-term care. The European Centre for Social Welfare Policy and Research. 2009. URL: https://www.euro.centre.org/publications/detail/385 [accessed 2021-07-01]
13. Rogers E. Diffusion of innovations. New York: Simon & Schuster; 2003.
14. Hartley J. New development: Eight and a half propositions to stimulate frugal innovation. Public Money & Management 2014 Apr 07;34(3):227-232 [FREE Full text] [doi: 10.1080/09540962.2014.908034]
15. Fuglsang L, Sørensen F. The balance between bricolage and innovation: management dilemmas in sustainable public innovation. The Service Industries Journal 2011 Mar;31(4):581-595 [FREE Full text] [doi: 10.1080/02642069.2010.504302]
16. Star SL, Ruhleder K. Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces. Information Systems Research 1996 Mar;7(1):111-134. [doi: 10.1287/isre.7.1.111]

17. Greenhalgh T, Procter R, Wherton J, Sugarhood P, Shaw S. The organising vision for telehealth and telecare: discourse analysis. BMJ Open 2012;2(4):e001574 [FREE Full text] [doi: 10.1136/bmjopen-2012-001574] [Medline: 22815469]
18. Boogerd EA, Arts T, Engelen LJ, van de Belt TH. "What Is eHealth": Time for An Update? JMIR Res Protoc 2015 Mar 12;4(1):e29 [FREE Full text] [doi: 10.2196/resprot.4065] [Medline: 25768939]
19. Ministry of Health and Care Services. Innovation in the Care Services. 2011. URL: https://www.regjeringen.no/contentassets/5fd24706b4474177bec0938582e3964a/en-gb/pdfs/nou201120110011000en_pdfs.pdf [accessed 2021-07-01]
20. The Ministry of Health and Care Services. Future Care. URL: https://www.regjeringen.no/contentassets/34c8183cc5cd43e2bd341e34e326dbd8/en-gb/pdfs/stm201220130029000engpdfs.pdf [accessed 2021-07-05]
21. The Norwegian Directorate of Health. Fagrapport om implementering av velferdsteknologi i de kommunale helse- og omsorgstjenestene 2013-2030 [Welfare technology. Professional report on the implementation of welfare technology in the municipal health and care services 2013-2030]. URL: https://www.helsedirektoratet.no/rapporter/implementering-av-velferdsteknologi-i-de-kommunale-helse-og-omsorgstjenestene-2013-2030/Implementering%20av%20velferdsteknologi%20i%20de%20kommunale%20helse-og%20omsorgstjenestene%202013-2030.pdf/_/attachment/inline/cf340308-0cb8-4a88-a6d7-4754ef126db9:6f3a196c2d353a9ef04c772f7cc0a2cb9d955087/Implementering%20av%20velferdsteknologi%20i%20de%20kommunale%20helse-og%20omsorgstjenestene%202013-2030.pdf [accessed 2021-07-05]
22. Lundvall B. Innovation studies: A personal interpretation of the state of the art. In: Fagerberg J, Martin B, Andersen E, editors. Innovation studies: Evolution and future challenges. Oxford: Oxford University Press; 2013.
23. Brevik K. A National Welfare Technology Programme in Norway, why and how? Norwegian Directorate of Health. 2014. URL: https://www.stjornarradid.is/media/velferdarraduneyti-media/media/veltek2014/Kristine----A-National-Welfare-Technology-Program-Norway1.pdf [accessed 2021-07-05]
24. The Welfare Technology Program. The Norwegian Association of Local and Regional Authorities (KS). 2018. URL: https://www.ks.no/fagomrader/helse-og-omsorg/velferdsteknologi3/nasjonalt-velferdsteknologiprogram/ [accessed 2021-07-05]
25. About the National Welfare Technology Program. The Norwegian Directorate of Health. 2019. URL: https://www.helsedirektoratet.no/tema/velferdsteknologi/velferdsteknologi [accessed 2021-07-05]
26. Kaiserfeld T. Innovation Monomania. In: Kaiserfeld T, editor. Beyond Innovation: Technology, Institution and Change as Categories for Social Analysis. London: Palgrave MacMillian; 2015.
27. Porter ME. What is value in health care? N Engl J Med 2010 Dec 23;363(26):2477-2481. [doi: 10.1056/NEJMp1011024] [Medline: 21142528]
28. Melting J. Andre gevinstrealiseringsrapport med anbefalinger. Nasjonalt velferdsteknologiprogram [Second report of realization of gains. The national welfare technology program]. Kommunesektorens organisasjon [The Norwegian Association of Local and Regional Authorities]. URL: https://www.ks.no/globalassets/1.0-andre-gevinstrealiseringsrapport-med-anbefalinger.pdf [accessed 2021-07-01]
29. Melting J, Frantzen L. Første gevinstrealiseringsrapport med anbefalinger. Nasjonalt velferdsteknologiprogram [First report of realization of gains and recommendations. The national welfare technology program]. Kommunesektorens organisasjon [The Norwegian Association of Local and Regional Authorities]. 2015. URL: https://www.helsedirektoratet.no/rapporter/gevinstrealiseringsrapporter-nasjonalt-velferdsteknologiprogram/F%C3%B8rste%20gevinstrealiseringsrapport%20%E2%80%93%20Nasjonalt%20velferdsteknologiprogram.pdf/_/attachment/inline/c879e5ef-715a-4277-9e89-e88f81e3b862:bbe6388b148899c2855be38afc2e163ce8065740/F%C3%B8rste%20gevinstrealiseringsrapport%20%E2%80%93%20Nasjonalt%20velferdsteknologiprogram.pdf [accessed 2021-07-01]
30. Andreassen HK, Kjekshus LE, Tjora A. Survival of the project: a case study of ICT innovation in health care. Soc Sci Med 2015 May;132:62-69. [doi: 10.1016/j.socscimed.2015.03.016] [Medline: 25795426]
31. Essén A, Lindblad S. Innovation as emergence in healthcare: unpacking change from within. Soc Sci Med 2013 Sep;93:203-211 [FREE Full text] [doi: 10.1016/j.socscimed.2012.08.035] [Medline: 23021848]
32. Halford S, Lotherington A, Obstfelder A, Dyb K. Getting the whole picture? New information and communication technologies in healthcare work and organization. Information, Communication & Society 2010 Apr;13(3):442-465 [FREE Full text] [doi: 10.1080/13691180903095856]
33. Barlow J, Hendy J. Adopting integrated mainstream telecare services. Lessons from the UK. Eurohealth 2009;15(1):8-10 [FREE Full text]
34. Pols J. Good relations with technology: Empirical ethics and aesthetics in care. Nurs Philos 2017 Jan;18(1):e12154. [doi: 10.1111/nup.12154] [Medline: 27758077]
35. Pols J, Willems D. Innovation and evaluation: taming and unleashing telecare technology. Sociol Health Illn 2011 Mar;33(3):484-498 [FREE Full text] [doi: 10.1111/j.1467-9566.2010.01293.x] [Medline: 21241338]
36. Bijker W, Hughes T, Pinch T. The social construction of technological systems: New directions in the sociology and history of technology. Cambridge: The MIT press; 2012.
37. Nicolini D. Practice theory, work, and organization: An introduction. Oxford: Oxford University Press; 2012.
38. Stokke R. From “Plug” to “play”: Making established technology innovations work in caring services. Lillehammer: The Inland Norway University of Applied Sciences; 2018.

39. Fagerberg J, Fosaas M, Sapprasert K. Innovation: Exploring the knowledge base. Research Policy 2012 Sep;41(7):1132-1153 [FREE Full text] [doi: 10.1016/j.respol.2012.03.008]
40. Fagerberg J, Martin MB, Andersen ES. Innovation Studies: Evolution and Future Challenges. Oxford: Oxford University Press; 2013.
41. Berge MS. Telecare acceptance as sticky entrapment: a realist review. Gerontechnology 2016;15(2):98-108. [doi: 10.4017/gt.2016.15.2.023.00]
42. Greenhalgh T, Procter R, Wherton J, Sugarhood P, Hinder S, Rouncefield M. What is quality in assisted living technology? The ARCHIE framework for effective telehealth and telecare services. BMC Med 2015;13:1-15 [FREE Full text] [doi: 10.1186/s12916-015-0279-6] [Medline: 25902803]
43. Procter R, Wherton J, Greenhalgh T, Sugarhood P, Rouncefield M, Hinder S. Telecare Call Centre Work and Ageing in Place. Comput Supported Coop Work 2016 Jan 15;25(1):79-105. [doi: 10.1007/s10606-015-9242-5]
44. Sánchez-Criado T, López D, Roberts C, Domènech M. Installing Telecare, Installing Users: Felicity Conditions for the Instauration of Usership. Science, Technology, & Human Values 2014 Jan 09;39(5):694-719 [FREE Full text] [doi: 10.1177/0162243913517011]
45. Benington J, Moore M. Public value in complex and changing times. In: Benington J, Moore M, editors. Public value: Theory and practice. London: Palgrave MacMillian; 2011:1-20.
46. Fuglesang L, Rønning R. Introduction: Framing innovation in public service sectors: A contextual approach. In: Fuglesang L, Rønning R, Enquist B, editors. Framing innovation in public service sectors. Oxfordshire: Routledge; 2014.
47. Rittel HWJ, Webber MM. Dilemmas in a general theory of planning. Policy Sci 1973 Jun;4(2):155-169. [doi: 10.1007/bf01405730]
48. Roberts N. Wicked problems and network approaches to resolution. International public management review 2000;1(1):1-19 [FREE Full text]
49. Greenhalgh T, Wherton J, Shaw S, Papoutsi C, Vijayaraghavan S, Stones R. Infrastructure Revisited: An Ethnographic Case Study of how Health Information Infrastructure Shapes and Constrains Technological Innovation. J Med Internet Res 2019 Dec 19;21(12):e16093 [FREE Full text] [doi: 10.2196/16093] [Medline: 31855184]
50. Denzin NK. The Research Act: A Theoretical Introduction to Sociological Methods. Piscataway: AldineTransaction; 1970.
51. The Norwegian Government. Reform of local government. The Norwegian Government. URL: https://www.regjeringen.no/no/tema/kommuner-og-regioner/kommunereform/reform-of-local-government/id2548429/ [accessed 2021-07-01]
52. Statistics Norway. Classification of centrality. URL: https://www.ssb.no/en/klass/klassifikasjoner/128/koder [accessed 2021-07-01]
53. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
54. Haugstveit FV, Otnes B. Vesentlig mer bruk av omsorgstjenester ved passerte 85 år [Significantly more use of care services after the age of 85]. 2019. URL: https://www.ssb.no/helse/artikler-og-publikasjoner/vesentlig-mer-bruk-av-omsorgstjenester-ved-passerte-85-ar [accessed 2021-07-01]
55. The Norwegian Directorate of Health. Anbefalinger om velferdsteknologiske løsninger i kommunene [Recommendations concerning welfare technology solutions in the municiplaites]. URL: https://www.helsedirektoratet.no/tema/velferdsteknologi/anbefalinger-om-velferdsteknologiske-losninger-i-kommunene [accessed 2021-07-01]
56. Stokke R. The Personal Emergency Response System as a Technology Innovation in Primary Health Care Services: An Integrative Review. J Med Internet Res 2016 Jul 14;18(7):e187 [FREE Full text] [doi: 10.2196/jmir.5727] [Medline: 27417422]
57. Høiland G, Willumsen E. Resistance-driven Innovation? Frontline Public Welfare Workers? Coping with Top-down Implementation. Nordic Journal of Working Life Studies 2018 Jun 04;8(2):23-42 [FREE Full text] [doi: 10.18291/njwls.v8i2.106153]
58. Kjær AM. Governance. Cambridge: Polity; 2004.
59. Moore M, Hartley J. Innovations in governance. Public Management Review 2008 Jan;10(1):3-20. [doi: 10.1080/14719030701763161]
60. Magnussen S. Innovativ samstyring mellom stat og kommuner: Fremforhandlet enighet eller frivillig tvang? En casestudie om statlige stimuleringsmidlers betydning ved utvikling av interkommunale lokamedisinske sentre [Innovative co-management between the state and municipalities: Negotiated agreement or voluntary coersion? A case study on the importance of state stimulants in the development of inter-municipal local medical centers]. Nordiske Organisasjonsstudier [The Nordic Journal of Organisation Studies] 2016;18(3):31-51.
61. Johannessen TB, Storm M, Holm AL. Safety for older adults using telecare: Perceptions of homecare professionals. Nurs Open 2019 Jul;6(3):1254-1261 [FREE Full text] [doi: 10.1002/nop2.328] [Medline: 31367452]
62. Linskell J, Dewsbury G. Assisted Living. In: Cowan D, Najafi L, editors. Handbook of Electronic Assistive Technology. Cambridge: Academic Press; 2019.
63. Slettebø A, Kirkevold M, Andersen B, Pedersen R, Halvorsen K, Nordhaug M, et al. Clinical prioritizations and contextual constraints in nursing homes--a qualitative study. Scand J Caring Sci 2010 Sep;24(3):533-540. [doi: 10.1111/j.1471-6712.2009.00745.x] [Medline: 20409056]

64. The Ministry of Health and Care Services. A full life - all your life — A Quality Reform for Older Persons. 2018. URL: https://www.helsedirektoratet.no/tema/leve-hele-livet-kvalitetsreformen-for-eldre/St%20Meld%2015%20-%20engelsk.pdf/_/attachment/inline/8561e891-57dd-447b-a184-895b739d74ce:c3c28dd80df596a4c8a05ffe6e5cf2e291198ba0/St%20Meld%2015%20-%20engelsk.pdf [accessed 2021-07-01]
65. Førland O, Ambugo E, Døhl, Folkestad B, Rostad H, Sundsbø A. Variasjon i kvalitet i omsorgstjenestene [Variation in quality of care services]. 2020. URL: https://omsorgsforskning.brage.unit.no/omsorgsforskning-xmlui/bitstream/handle/11250/2658156/Forskningsdel%3b%20rapport_01_2020%20facing-web.pdf?sequence=1&isAllowed=y [accessed 2021-07-01]
66. Mossberger K, Tolbert CJ, Franko W. The Need for Urban Broadband Policy. In: Mossberger K, Tolbert CJ, Franko W, editors. Digital Cities: The Internet and the Geography of Opportunity. Oxford: Oxford University Press; 2013.
67. Glaeser E. Triumph of the City: How Our Greatest Invention Makes Us Richer, Smarter, Greener, Healthier, and Happie. London: Penguin Books; 2012.
68. Postema TRF, Peeters JM, Friele RD. Key factors influencing the implementation success of a home telecare application. Int J Med Inform 2012 Jun;81(6):415-423 [FREE Full text] [doi: 10.1016/j.ijmedinf.2011.12.003] [Medline: 22226925]
69. National Center for Chronic Disease Prevention and Health Promotion. Telehealth in Rural Communities. 2019. URL: https://www.cdc.gov/chronicdisease/resources/publications/factsheets/telehealth-in-rural-communities.htm [accessed 2021-07-01]
70. Moffatt JJ, Eley DS. The reported benefits of telehealth for rural Australians. Aust Health Rev 2010 Aug;34(3):276-281. [doi: 10.1071/AH09794] [Medline: 20797357]
71. PA Consulting Group. Organisering og dimensjonering av responssentertjeneste [Organization and dimensioning of response center service]. 2016. URL: https://www.helsedirektoratet.no/tema/velferdsteknologi/rapporter-og-utredninger/Organisering%20og%20dimensjonering%20av%20responssentertjeneste.pdf/_/attachment/inline/20f3e1c3-d3ba-4dff-b403-233767108027:93d4afbea8a6c7e370ffc7f8b6cac19cf01f2da3/Organisering%20og%20dimensjonering%20av%20responssentertjeneste.pdf [accessed 2021-07-01]
72. Pols J, Moser I. Cold technologies versus warm care? On affective and social relations with and through care technologies. Alter 2009 Apr;3(2):159-178 [FREE Full text] [doi: 10.1016/j.alter.2009.01.003]
73. Isaksen J, Stokke R. Utfordringer med velferdsteknologibegrepet [Challenges with the concept of welfare technology]. Tidsskrift for omsorgsforskning [Journal of Care Reseach] 2017;3(2):91-94. [doi: 10.18261/issn.2387-5984-2017-02-06]
74. Allen D. The Invisible Work of Nurses. Hospitals, Organisation and Healthcare. London: Routledge; 2014.
75. Hartley J, Rashman L. How is knowledge transferred between organizations involved in change. In: Wallace M, Fertig M, Schneller E, editors. Managing change in the public services. Hoboken: Wiley; 2007:173-192.
76. Norwegian Institute of Public Health. Health among the elderly in Norway. 2016. URL: https://www.fhi.no/en/op/hin/groups/health-and-ageing-in-norway/ [accessed 2021-07-01]
77. Wyatt JC, Sullivan F. eHealth and the future: promise or peril? BMJ 2005 Dec 10;331(7529):1391-1393 [FREE Full text] [doi: 10.1136/bmj.331.7529.1391] [Medline: 16339252]
78. Tongur S, Engwall M. Exploring window of opportunity dynamics in infrastructure transformation. Environmental Innovation and Societal Transitions 2017 Dec;25:82-93. [doi: 10.1016/j.eist.2016.12.003]
79. Toivonen M. Different types of innovation processes in services and their organisational implications. In: Gallouj F, Djellal F, editors. The handbook of innovation and services. A Multi-disciplinary Perspective. Cheltenham Glos: Edward Elgar Publishing Limited; 2010:A.
80. Geels FW. Technological transitions as evolutionary reconfiguration processes: a multi-level perspective and a case-study. Research Policy 2002 Dec;31(8-9):1257-1274. [doi: 10.1016/s0048-7333(02)00062-8]
81. Barrutia JM, Echebarria C. Drivers of exploitative and explorative innovation in a collaborative public-sector context. Public Management Review 2018 Aug 28;21(3):446-472. [doi: 10.1080/14719037.2018.1500630]
82. Torfing J. Collaborative innovation in the public sector: the argument. Public Management Review 2018 Feb 02;21(1):1-11 [FREE Full text] [doi: 10.1080/14719037.2018.1430248]
83. Wiberg U, Limani I. Intermunicipal collaboration: a smart alternative for small municipalities? Scandinavian Journal of Public Administration 2015;19(1):63-82 [FREE Full text]
84. Kannampallil TG, Schauer GF, Cohen T, Patel VL. Considering complexity in healthcare systems. J Biomed Inform 2011 Dec;44(6):943-947 [FREE Full text] [doi: 10.1016/j.jbi.2011.06.006] [Medline: 21763459]
85. van den Broek J, Boselie P, Paauwe J. Multiple Institutional Logics in Health Care: 'Productive Ward: Releasing Time to Care'. Public Management Review 2013 Apr 18;16(1):1-20 [FREE Full text] [doi: 10.1080/14719037.2013.770059]
86. Geels FW, Schot J. Typology of sociotechnical transition pathways. Research Policy 2007 Apr;36(3):399-417. [doi: 10.1016/j.respol.2007.01.003]
87. Geels FW. Micro-foundations of the multi-level perspective on socio-technical transitions: developing a multi-dimensional model of agency through crossovers between social constructivism, evolutionary economics and neo-institutional theory. Technol Forecast Soc Change 2020 Mar;152:119894. [doi: 10.1016/j.techfore.2019.119894]
88. Kjelvik J, Jønsberg E. Botid i sykehjem og varighet av tjenester til hjemmeboende [Living-time in nursing homes and duration of services for home residents]. 2017. URL: https://www.helsedirektoratet.no/rapporter/

botid-i-sykehjem-og-varighet-av-tjenester-til-hjemmeboende/2017-02%20Botid%20i%20sykehjem%20og%20varighet%20av%20tjenester%20til%20hjemmeboende.pdf/_/attachment/inline/9f8fa68c-5969-4147-95d1-2177464084de:8a6b1b6e741b917894778a5ef81610764635ea4c/2017-02%20Botid%20i%20sykehjem%20og%20varighet%20av%20tjenester%20til%20hjemmeboende.pdf [accessed 2021-07-01]

89. Sherwin S, Winsby M. A relational perspective on autonomy for older adults residing in nursing homes. Health Expect 2011 Jun;14(2):182-190 [FREE Full text] [doi: 10.1111/j.1369-7625.2010.00638.x] [Medline: 21029285]
90. Koren MJ. Person-centered care for nursing home residents: the culture-change movement. Health Aff 2010 Feb;29(2):312-317. [doi: 10.1377/hlthaff.2009.0966] [Medline: 20056692]
91. Thornton P, Ocasio W, Lounsbury M. The Institutional Logics Perspective: A New Approach to Culture, Structure and Process. Oxford: Oxford University Press; 2012.

Edited by R Kukafka; submitted 08.07.20; peer-reviewed by D Trupia, S Levy, K Baudin; comments to author 11.09.20; revised version received 19.10.20; accepted 24.05.21; published 16.08.21.

Please cite as:
Rostad HM, Stokke R
Integrating Welfare Technology in Long-term Care Services: Nationwide Cross-sectional Survey Study
J Med Internet Res 2021;23(8):e22316
URL: https://www.jmir.org/2021/8/e22316
doi:10.2196/22316
PMID:34398791

©Hanne Marie Rostad, Randi Stokke. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 16.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Digital Orientation of Health Systems in the Post–COVID-19 “New Normal” in the United States: Cross-sectional Survey

Jiban Khuntia[1], PhD; Xue Ning[1], PhD; Rulon Stacey[1], PhD

CU Business School, University of Colorado Denver, Denver, CO, United States

Corresponding Author:
Jiban Khuntia, PhD
CU Business School
University of Colorado Denver
1475 Lawrence Street
Denver, CO, 80202
United States
Phone: 1 3038548024
Email: jiban.khuntia@ucdenver.edu

Abstract

Background: Almost all health systems have developed some form of customer-facing digital technologies and have worked to align these systems to their existing electronic health records to accommodate the surge in remote and virtual care deliveries during the COVID-19 pandemic. Others have developed analytics-driven decision-making capabilities. However, it is not clear how health systems in the United States are embracing digital technologies and there is a gap in health systems’ abilities to integrate workflows with expanding technologies to spur innovation and futuristic growth. There is a lack of reliable and reported estimates of the current and futuristic digital orientations of health systems. Periodic assessments will provide imperatives to policy formulation and align efforts to yield the transformative power of emerging digital technologies.

Objective: The aim of this study was to explore and examine differences in US health systems with respect to digital orientations in the post–COVID-19 “new normal” in 2021. Differences were assessed in four dimensions: (1) analytics-oriented digital technologies (AODT), (2) customer-oriented digital technologies (CODT), (3) growth and innovation–oriented digital technologies (GODT), and (4) futuristic and experimental digital technologies (FEDT). The former two dimensions are foundational to health systems’ digital orientation, whereas the latter two will prepare for future disruptions.

Methods: We surveyed a robust group of health system chief executive officers (CEOs) across the United States from February to March 2021. Among the 625 CEOs, 135 (22%) responded to our survey. We considered the above four broad digital technology orientations, which were ratified with expert consensus. Secondary data were collected from the Agency for Healthcare Research and Quality Hospital Compendium, leading to a matched usable dataset of 124 health systems for analysis. We examined the relationship of adopting the four digital orientations to specific hospital characteristics and earlier reported factors as barriers or facilitators to technology adoption.

Results: Health systems showed a lower level of CODT (mean 4.70) or GODT (mean 4.54) orientations compared with AODT (mean 5.03), and showed the lowest level of FEDT orientation (mean 4.31). The ordered logistic estimation results provided nuanced insights. Medium-sized ($P<.001$) health systems, major teaching health systems ($P<.001$), and systems with high-burden hospitals ($P<.001$) appear to be doing worse with respect to AODT orientations, raising some concerns. Health systems of medium ($P<.001$) and large ($P=.02$) sizes, major teaching health systems ($P=.07$), those with a high revenue ($P=.05$), and systems with high-burden hospitals ($P<.001$) have less CODT orientation. Health systems in the midwest ($P=.05$) and southern ($P=.04$) states are more likely to adopt GODT, whereas high-revenue ($P=.004$) and investor-ownership ($P=.01$) health systems are deterred from GODT. Health systems of a medium size, and those that are in the midwest ($P<.001$), south ($P<.001$), and west ($P=.01$) are more adept to FEDT, whereas medium ($P<.001$) and high-revenue ($P<.001$) health systems, and those with a high discharge rate ($P=.04$) or high burden ($P=.003$, $P=.005$) have subdued FEDT orientations.

Conclusions: Almost all health systems have some current foundational digital technological orientations to glean intelligence or service delivery to customers, with some notable exceptions. Comparatively, fewer health systems have growth or futuristic digital orientations. The transformative power of digital technologies can only be leveraged by adopting futuristic digital technologies. Thus, the disparities across these orientations suggest that a holistic, consistent, and well-articulated direction across the United States remains elusive. Accordingly, we suggest that a policy strategy and financial incentives are necessary to spur a well-visioned and articulated digital orientation for all health systems across the United States. In the absence of such a policy to collectively leverage digital transformations, differences in care across the country will continue to be a concern.

(J Med Internet Res 2021;23(8):e30453) doi:10.2196/30453

KEYWORDS

post–COVID-19; digital orientation; health systems; digital transformation; digital health; telehealth; telemedicine; COVID-19; impact; insight; cross-sectional; survey; United States; electronic health record; EHR

Introduction

Background

The COVID-19 pandemic aggravated the perennial issues of cost, quality, and delivery challenges of health care in the United States. Simultaneously, the COVID-19 pandemic has also opened up newer directions to solve some of these challenges. Digital technologies have come to the forefront in solving many of the challenges. The critical role of technology in fighting the pandemic through effective tracking of the virus across the world is undeniable. Health systems have used existing health record systems along with surveillance and monitoring applications to gather, collate, analyze, and present information to the government to make meaningful and valuable decisions to help in the pandemic. Several technologies played an essential role in informing health systems and frontline health professionals to fight the crisis.

The scope of health information technologies has traditionally been limited to electronic health or medical records, with sporadic examples of data and intelligence-based decision-making using the recorded data. Digital records have been touted to increase the potential to improve health care providers' efficiency and effectiveness. Key functionalities of electronic health records such as computerized provider order entry for medications, electronic prescribing, or using clinical decision support systems have helped to achieve some of the objectives [1].

The last decade of policies around health information technologies has propelled the adoption of digital records to almost 100%, although the comprehensiveness and meaningful use of these records are still debatable [2]. Nevertheless, the scope of electronic health records to play a transformative role in health care is limited. The fact remains that some health systems are still using a set of basic functionalities rather than fully leveraging more comprehensive functionalities. The result is that these health systems cannot fully shape future decisions and strategies for health care [3].

Industry sectors other than health care have leveraged the transformative potential of digital technologies beyond the capture, archiving, and use of data using digital records only. The increasing prevalence of digital technology is fundamentally transforming how businesses create value. Research has postulated that technology and strategy align together to drive proper digital transformation and ultimately provide a competitive advantage.

In this context, we sought to address the following questions: (1) What are the digital orientations of health systems in the post COVID-2019 new normal? (2) How can such orientations be measured and compared across health systems to provide a systemic evaluation across the United States? (3) What are the factors that may influence the digital orientations of health systems?

With anticipated health care spending to reach US $5.7 trillion by 2026, the time is right to enhance public policy understanding of digital technologies and to build the strategic imperatives around those technologies. However, to change the current standard practices and install improved digital-based technologies into health systems while leveraging these improvements throughout the industry, a comprehensive assessment of the health systems' digital orientations in 2021 is needed to inform research and practice.

Literature Review: Digital Orientations

Digital orientation as a strategic direction exhibits superior performance by using and leveraging technology in different ways and through different means while maintaining a view of current and futuristic options [4]. A specific digital orientation can be examined from perspectives such as technology scope and capabilities [4]. Different digital orientations shape the way organizations create and adapt behaviors and resources [5,6], similar to market orientation (eg, [7]) and entrepreneurial orientation (eg, [8]), which have been extensively studied as sources of competitive advantage [9].

Digital orientation with a strategic and futuristic direction will nurture and implement subsequent digitalization initiatives. As a result, digital orientation will create value beyond what is seen as the immediate returns of digital investments [10] and direct to an unprecedented scope and degree of openness, driven by generative and unpredictable processes and contingent on the specific affordances of digital technologies as realized in other sectors [11]. Digital orientation will change the traditional competitive logic to stimulate distinct and novel processes, and managerial and organizational alignment.

The basic step of digital orientation starts with the technologies that support the existing functions of an organization on a day-to-day basis [12,13]. Given that electronic health records have been well disseminated in US health care, the data-driven clinical and administrative decision-making based on mining applications and tools to analyze the data captured along with data available from the records encompass such an orientation. We coin this category of basic orientation as *analytics orientation.*

Another category is *customer-oriented digital technologies* (CODT), which involve technical interfaces through which customers can access services that enable standardized delivery of services [14,15] to provide increased flexibility of access

[16,17]. Mobile technologies are an example of tools that provide such access. Social media–integrated tools and applications result in different avenues for the customer to reach these services [18,19].

A set of emerging technologies helps health systems to foster *growth and innovation orientation*. These technologies help to reevaluate and reengineer several business functions, similar to the enterprise resource planning applications [20]. The underlying concept for this orientation is to be innovative in changing business functions and processes, and extending these innovations across partnering businesses to change the value chain. For instance, information exchange with organizations helps to provide just-in-time care effectively while extending care provisions across health systems [21]. Similarly, virtual and remote care models require that the diagnosis and treatment involving physician and patient interactions be redesigned and aligned to newer value-based models relative to the earlier fee-based models.

Finally, *futuristic and experimental digital technologies* (FEDT) are being trialed or experimented with in terms of their potential to change the practice and delivery of health care [22]. These may not be widely disseminated, and the value may not be predictably assessed as is the case for the growth-oriented technologies. Examples of this category would include robotics applications, wearable chips, and tracking devices [23]. Artificial intelligence and machine learning applications are also being introduced to health care, with some value potential, but are waiting for broader dissemination [24].

Delineating the current stage of the four digital orientations described above will aid in guiding strategies and policies in health care. The US health care systems need an overarching digitally enabled strategic orientation to holistically mirror the heightened, transfunctional role of digitalization across the sector. Assessment of these four orientations in 2021 is a first step to guide future actions in this direction.

Study Aims

Accordingly, the aim of this study was to explore and measure the differences between these four types of orientations in health systems with different characteristics, including size, region, ownership status, teaching status, revenue, number of physicians, hospitals, and other factors. We also explored the impacts of these factors on the levels of the four digital orientations. This study is unique in focusing on the digital orientation of system-wide differences across health systems in the United States. The findings of this study will provide implications for the strategic development of health systems in the post COVID-19 era, and suggest a top-level US health systems digital strategy and plan that can shape the development blueprints for all health systems and the nation.

Methods

Data Collection

The effort to assess the digital orientations of health systems is part of a broad project on the climate of health systems undertaken by the Health Administration Research Consortium at the Business School of the University of Colorado Denver. The idea of monitoring health systems emerged from our observations and conversations with several chief executive officers (CEOs) of health systems during the COVID-19 pandemic. The objective was to collect and inform the insights of health systems’ CEOs that will help inform policymakers, practitioners, and academic stakeholders as they collaborate to create ongoing strategies to help the industry respond to this pandemic and prepare for the next crisis.

For the inaugural climate study, a survey questionnaire was developed in December 2020 to collect data from the health systems and scientifically study the climate that health systems face. The survey items were derived from the prior literature with rewording of the questions to fit the health systems context. Inputs were taken from researchers, consultants, and executives with expertise to design the questions. The survey was validated using a scientific process of experts’ evaluations and was pilot-tested with five top executives who are part of the Health Administration Program Advisory Board. The survey questionnaire was revised and finalized in January 2021 [25].

We compiled a contacts list of CEOs of a total of 624 health systems across the United States using information from multiple sources, contacts, professional collections, websites, and annual reports. We mounted the survey instrument on a professional survey platform. We mapped the emails to the platform to create unique trackable links for each health system. We sent the invitation and solicitation emails to the CEOs in multiple rounds between January 25 and March 2, 2021. Along with this, the authors called several CEOs and solicited completion of the survey instrument either online or in paper format. The researchers also requested CEOs who had participated in the survey to share the link with other CEO colleagues. We received a total of 148 responses, with a 24% response rate. We could not use 13 incomplete responses, leaving 135 final usable responses for analysis.

The size of the 135 health systems represented in this survey varied from 1 to 18 hospitals with 176 to 75,000 employees. The annual revenue in 2020 of the health systems ranged from US $0.7 million to US $14 billion. The health systems aggregately represent US $0.3 trillion in revenues with 1.1 million employees across the United States.

We then matched the survey dataset with the secondary data collected from the Agency for Healthcare Research and Quality Hospital Compendium to obtain a better and more complete picture of the health systems. Finally, we obtained a dataset of 124 health systems that are located across the United States. We analyzed this combined dataset to report several insights in this study.

Variables and Measures

Table 1 shows the description of the variables used in this study. The primary dependent variables were the four digital orientations: (1) analytics-oriented digital technologies (AODT), (2) CODT, (3) growth and innovation–oriented digital technologies (GODT), and (4) FEDT.

Table 1. Description of variables, including survey questions and coding scheme.

Variable	Description
Survey question: “To what extent do you consider the following digital technologies to be important and creating value for your health systems?”[a]	
AODT	Analytics-oriented digital technologies, including data mining and analysis, data-driven administrative decision-making, and data-driven clinical decision-making (Cronbach α=.77)
CODT	Customer-oriented digital technologies, including mobile technologies for customer engagement (eg, social media tools, applications, and integration), rendering higher customer-oriented services (Cronbach α=.57)
GODT	Growth and innovation–oriented digital technologies, including reengineering several business functions, providing innovation potential within the organization, and providing innovation capacity in collaboration with external organizations (Cronbach α=.57)
FEDT	Futuristic and experimental digital technologies, including virtual monitoring using wearables, chips, and tracking devices; robotics applications in health care; and artificial intelligence and machine learning (Cronbach α=.91)
Contingent variables	
SIZE	Total beds managed by the health system across all hospitals, reported by the Agency for Healthcare Research and Quality (AHRQ) Hospital Compendium
SIZE_B-SMALL	<100 beds
SIZE_B-MEDIUM	100-400 beds
SIZE_B-LARGE	>400 beds
REGION	Primary location of the health system in the United States, following the Census Bureau categorization
REGION-NE	Northeast
REGION-MW	Midwest
REGION-SOUTH	South
REGION-WEST	West
TEACHING	Teaching status of a health system
TEACHING-NON	nonteaching
TEACHING-MINOR	minor teaching
TEACHING-MAJOR	major teaching
REVENUE	Annual revenue (US $) of the health system across all hospitals
REVENUE-LOW	<2 billion
REVENUE-MEDIUM	2-5 billion
REVENUE-HIGH	>5 billion
HIGH-DSH-HOSP	The health system includes at least one Disproportionate Share Hospital (DSH) (1=yes, 0=no), which refers to a hospital that serves a significantly disproportionate number of low-income patients and receives payments from the Centers for Medicaid and Medicare Services of the United States to cover the costs of providing care to uninsured patients.
BURDEN	This reflected the uncompensated care, as an overall measure of hospital care provided for which no payment was received from the patient or insurer.
HIGH-BURDEN-SYS	Health system–wide uncompensated care burden flag (1=yes, 0=no)
HIGH-BURDEN-HOSP	The health system includes at least one high uncompensated care burden hospital (1=yes, 0=no)
OWNERSHIP	Predominantly investor-owned hospitals (1=yes, 0=no)
PHYSICIANS	Number of physicians in the health system as reported by the AHRQ Hospital Compendium.
HOSPITALS	Number of hospitals in the health system as reported by the AHRQ Hospital Compendium

[a] All questions were measured using a 7-point Likert scale (1=strongly disagree to 7=strongly agree).

These four variables were measured based on a 7-point Likert scale of relevant items. The reliability was tested using Cronbach α, which was higher than the acceptable threshold of .50 for reflective items and measures used in this study [26].

The independent variables of this study covered several categories: size, region, teaching status, revenue, and other system characteristics. These variables were coded to reflect the characteristics of a health system, which may influence its digital orientations (see Table 1). In summary, three size variables were used to measure the number of beds across a health system (SIZE_B-SMALL, SIZE_B-MEDIUM, SIZE_B-LARGE); four region variables were used to reflect the location of a health system (REGION-NE, REGION-MW, REGION-SOUTH, REGION-WEST), there were three teaching status–related variables (TEACHING-NON, TEACHING-MINOR, TEACHING-MAJOR), and three revenue variables were used to measure the annual revenue of a health system (REVENUE-LOW, REVENUE-MEDIUM, REVENUE-HIGH). Other variables included those related to the existence of Disproportionate Share Hospital (DSH) patients (HIGH-DSH-HOSP), uncompensated care burden (HIGH-BURDEN-SYS and HIGH-BURDEN-HOSP), ownership status (OWNERSHIP), number of physicians (PHYSICIANS), and number of hospitals (HOSPITALS).

Sample Statistics

The descriptive statistics and pairwise correlations among the key variables used in this study are presented in Table 2 and Table 3. As shown in Table 2, health systems have a lower level of customer or growth orientations compared to AODT orientations, and showed the least amount of FEDT.

In addition, to make sure there is no nonresponse bias, we compared the characteristics of responding and nonresponding health systems. The detailed comparisons are provided in Table 4. The *t* test results for all comparisons indicated no significant difference between respondents and nonrespondents.

Table 2. Summary statistics (N=124).

Variable[a]	Mean (SD)	Minimum	Maximum
AODT[b]	5.03 (1.37)	2	6.67
CODT[c]	4.70 (1.35)	2.33	7
GODT[d]	4.54 (1.23)	2.33	7
FEDT[e]	4.31 (1.54)	1	7
SIZE_B-SMALL	0.09 (0.28)	0	1
SIZE_B-MEDIUM	0.37 (0.49)	0	1
SIZE_B-LARGE	0.54 (0.50)	0	1
REGION-NE	0.22 (0.42)	0	1
REGION-MW	0.24 (0.43)	0	1
REGION-SOUTH	0.35 (0.48)	0	1
REGION-WEST	0.18 (0.38)	0	1
TEACHING-NON	0.30 (0.46)	0	1
TEACHING-MINOR	0.48 (0.50)	0	1
TEACHING-MAJOR	0.22 (0.41)	0	1
REVENUE-LOW	0.61 (0.49)	0	1
REVENUE-MEDIUM	0.23 (0.43)	0	1
REVENUE-HIGH	0.15 (0.35)	0	1
HIGH-DSH-HOSP	0.33 (0.47)	0	1
HIGH-BURDEN-SYS	0.20 (0.40)	0	1
HIGH-BURDEN-HOSP	0.30 (0.46)	0	1
OWNERSHIP	0.02 (0.13)	0	1
PHYSICIANS	1.84 (0.80)	1	3
HOSPITALS	1.50 (0.77)	1	3

[a] See Table 1 for variable and code definitions.

[b] AODT: analytics-oriented digital technologies

[c] CODT: customer-oriented digital technologies.

[d] GODT: growth and innovation–oriented digital technologies.

[e] FEDT: futuristic and experimental digital technologies.

Table 3. Pairwise correlations (Pearson *r* and *P* values) among key variables (N=124).

Variable[a]	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. AODT[b]																			
r	1.00	0.70	0.06	–0.15	–0.13	0.03	–0.06	–0.02	0.16	–0.01	–0.11	0.04	–0.07	0.07	–0.03	–0.05	0.01	–0.05	0.06
P	—[c]	<.001	.52	.09	.14	.75	.50	.81	.08	.95	.23	.65	.41	.42	.75	.56	.88	.62	.51
2. CODT[d]																			
r	0.70	1.00	–0.13	0.01	–0.07	–0.05	–0.01	–0.16	0.24	–0.01	–0.15	–0.13	–0.02	–0.02	0.08	–0.14	0.01	–0.10	–0.04
P	<.001	—	.16	.94	.46	.58	.90	.07	.01	.91	.09	.15	.85	.82	.37	.11	.89	.16	.63
3. GODT[e]																			
r	0.06	–0.13	1.00	0.43	–0.08	0.13	0.13	0.03	0.001	–0.01	0.10	0.15	–0.02	–0.03	0.06	0.11	–0.21	0.08	0.08
P	.52	.16	—	<.001	.40	.15	.15	.71	.99	.92	.29	.09	.82	.77	.50	.23	.02	.37	.36
4. FEDT[f]																			
r	–0.15	0.01	0.43	1.00	0.01	0.06	0.10	–0.01	0.16	–0.08	0.07	0.003	0.0001	–0.15	0.22	–0.01	–0.15	0.04	–0.07
P	.09	.94	<.001	—	.92	.54	.28	.95	.07	.38	.44	.98	.99	.10	.01	.94	.09	.63	.41
5. SIZE_B-MED																			
r	–0.13	–0.07	–0.08	0.01	1.00	–0.83	–0.08	–0.01	–0.05	–0.10	–0.16	–0.27	–0.32	–0.08	–0.01	–0.21	0.03	–0.60	–0.46
P	.14	.46	.40	.92	—	<.001	.36	.90	.58	.29	.07	.003	<.001	.39	.89	.02	.71	<.001	<.001
6. SIZE_B-LARGE																			
r	0.03	–0.05	0.13	0.06	–0.83	1.00	–0.01	0.11	0.005	0.26	0.25	0.36	0.38	0.17	0.06	0.28	–0.14	0.75	0.54
P	.75	.58	.15	.54	<.001	—	.93	.23	.96	.003	.005	<.001	<.001	.06	.51	.001	.12	<.001	<.001
7. REGION-MW																			
r	–0.06	–0.01	0.13	0.10	–0.08	–0.01	1.00	–0.42	–0.26	–0.01	–0.12	–0.09	0.09	–0.20	–0.10	0.002	–0.07	0.02	0.02
P	.50	.90	.15	.28	.36	.93	—	<.001	.003	.91	.20	.32	.33	.03	.29	.98	.42	.83	.79
8. REGION-SOUTH																			
r	–0.02	–0.16	0.03	–0.01	–0.01	0.11	–0.42	1.00	–0.34	0.07	0.02	0.15	–0.02	0.12	0.17	0.22	0.17	0.002	0.04
P	.81	.07	.71	.95	.90	.23	<.001	—	<.001	.44	.85	.10	.84	.17	.05	.02	.06	.98	.63
9. REGION-WEST																			
r	0.16	0.24	0.001	0.16	–0.05	0.005	–0.26	–0.34	1.00	–0.02	–0.04	–0.06	–0.01	0.12	0.08	0.02	–0.06	–0.04	–0.03
P	.08	.01	.99	.07	.58	.96	.003	<.001	—	.83	.66	.53	.89	.18	.36	.82	.51	.67	.76
10. TEACHNG-MINOR																			
r	–0.01	–0.01	–0.01	–0.08	–0.10	0.26	–0.01	0.07	–0.02	1.00	–0.50	0.12	0.07	0.05	–0.12	0.08	0.01	0.19	0.26
P	.95	.91	.92	.38	.29	.003	.91	.44	.83	—	<.001	.18	.47	.57	.20	.35	.95	.03	.003
11. TEACHNG-MAJOR																			
r	–0.11	–0.15	0.10	0.07	–0.16	0.25	–0.12	0.02	–0.04	–0.50	1.00	0.12	0.17	0.34	0.03	0.13	–0.07	0.38	0.06
P	.23	.09	.29	.44	.07	.005	.20	.85	.66	<.001	—	.17	.06	<.001	.77	.16	.46	<.001	.48
12. REVENUE-MED																			
r	0.04	–0.13	0.15	0.003	–0.27	0.36	–0.09	0.15	–0.06	0.12	0.12	1.00	–0.23	–0.06	–0.04	0.14	–0.07	0.26	0.29
P	.65	.15	.09	.98	.003	<.001	.32	.10	.53	.18	.17	—	.01	.48	.66	.12	.44	.004	.001
13. REVENUE-HIGH																			
r	–0.07	–0.02	–0.02	0.0001	–0.32	0.38	0.09	–0.02	–0.01	0.07	0.17	–0.23	1.00	0.15	–0.04	–0.02	–0.05	0.51	0.30
P	.41	.85	.82	.99	<.001	<.001	.33	.84	.89	.47	.06	.01	—	.10	.69	.84	.56	<.001	<.001
14. HIGH-DSH-HOSP																			

Variable[a]	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
r	0.07	–0.02	–0.03	–0.15	–0.08	0.17	–0.20	0.12	0.12	0.05	0.34	–0.06	0.15	1.00	–0.01	0.18	0.05	0.23	0.17
P	.42	.82	.77	.10	.39	.06	.03	.17	.18	.57	<.001	.48	.10	—	.90	.05	.61	.01	.06
15. HIGH-BURD-SYS																			
r	–0.03	0.08	0.06	0.22	–0.01	0.06	–0.10	0.17	0.08	–0.12	0.03	–0.04	–0.04	–0.01	1.00	0.42	–0.06	–0.10	–0.20
P	.75	.37	.50	.01	.89	.51	.29	.05	.36	.20	.77	.66	.69	.90	—	<.001	.48	.27	.03
16. HIGH-BURD-HOSP																			
r	–0.05	–0.14	0.11	–0.01	–0.21	0.28	0.002	0.22	0.02	0.08	0.13	0.14	–0.02	0.18	0.42	1.00	–0.08	0.18	0.31
P	.56	.11	.23	.94	.02	.001	.98	.02	.82	.35	.16	.12	.84	.05	<.001	—	.36	.05	<.001
17. OWNERSHIP																			
r	0.01	0.01	–0.21	–0.15	0.03	–0.14	–0.07	0.17	–0.06	0.01	–0.07	–0.07	–0.05	0.05	–0.06	–0.08	1.00	0.18	0.31
P	.88	.89	.02	.09	.71	.12	.42	.06	.51	.95	.46	.44	.56	.61	.48	.36	—	.55	.36
18. PHYSICIANS																			
r	–0.05	–0.10	0.08	0.04	–0.60	0.75	0.02	0.002	–0.04	0.19	0.38	0.26	0.51	0.23	–0.10	0.18	–0.05	1.00	–0.08
P	.62	.16	.37	.63	<.001	<.001	.83	.98	.67	.03	<.001	.004	<.001	.01	.27	.05	.55	—	<.001
19. HOSPITALS																			
r	0.06	–0.04	0.08	–0.07	–0.46	0.54	0.02	0.04	–0.03	0.26	0.06	0.29	0.30	0.17	–0.20	0.31	–0.08	0.57	1.00
P	.51	.63	.36	.41	<.001	<.001	.79	.63	.76	.003	.48	.001	<.001	.06	.03	<.001	.36	<.001	—

[a] See Table 1 for variable and code definitions.

[b] Not applicable.

[c] AODT: analytics-oriented digital technologies

[d] CODT: customer-oriented digital technologies.

[e] GODT: growth and innovation–oriented digital technologies.

[f] FEDT: futuristic and experimental digital technologies.

Table 4. Characteristics of responding and nonresponding health systems.

Characteristics	Respondents, n (%) (N=124)	Nonrespondents, n (%) (N=511)	*t* value (*df*)
Size			
Small (6-99 beds)	11 (8.9)	43 (8.4)	–0.19 (50)
Medium (100-399 beds)	45 (36.3)	212 (41.5)	–0.56 (254)
Large (≥400 beds)	68 (54.9)	256 (50.1)	1.41 (325)
Region			
Northeast	27 (21.8)	118 (23.1)	0.07 (141)
Midwest	30 (24.2)	133 (26.0)	0.55 (162)
South	45 (36.3)	167 (32.7)	–0.48 (214)
West	22 (17.7)	93 (18.2)	–0.12 (112)
Physicians[a]			
Small (51-199 physicians)	50 (40.3)	189 (37.0)	–0.74 (238)
Medium (200-999 physicians)	41 (33.1)	204 (39.9)	–0.69 (243)
Large (≥1000 physicians)	33 (26.6)	118 (23.1)	1.53 (150)
Hospitals[a]			
Small (1-3 hospitals)	83 (66.9)	338 (66.1)	–1.27 (420)
Medium (4-6 hospitals)	20 (16.1)	66 (12.9)	–0.02 (85)
Large (≥7 hospitals)	21 (16.9)	107 (20.9)	0.81 (126)
Ownership status			
Investor-owned	3 (2.4)	15 (2.9)	–0.85 (15)
Noninvestor-owned	121 (97.6)	496 (97.1)	0.85 (616)
Teaching status			
Major teaching	29 (23.3)	138 (27.0)	–0.15 (186)
Minor teaching	58 (46.8)	225 (44.0)	–0.61 (280)
Nonteaching	37 (29.8)	148 (29.0)	0.85 (163)

[a] The numbers of physicians and hospitals are presented in this table in different categories for easy comparison across respondents and nonrespondents.

Statistical Analysis

We used ordered logit regressions to estimate the relationships of the four digital orientations to specific hospital characteristics. All four dependent variables are ordinal variables to drive the decision for ordered logit regressions. This approach does not assume equal intervals between levels in the dependent variable. The ordered logit model is as follows:

$$Y_i^* = \beta X_i + e_i$$

Where Y_i^* is the propensity of respondents to indicate higher levels of the four digital orientations (ie, AODT, CODT, GODT, FEDT), X_i is a set of explanatory variables, β a vector of parameters, and e_i represents disturbances.

Rather than observing Y_i^*, we observed the ordinal dependent variable Y_i depending on the values of thresholds or cut-off points τ_{m-1} and τ_m. The probability distribution of Y_i is given as follows:

$$\Pr(Y_i = m | \mathbf{X}_i = F(\tau_m - X\beta) - F(\tau_{m-1} - X\beta)$$

where

Results

Table 5 shows the results of the ordered logit model estimation. Each column presents the results for each of the four digital orientations.

First, we found that the two size variables have a significantly negative association with AODT. In particular, the medium size variable showed high statistical significance at $P<.001$. This suggests that smaller-sized health systems tend to adopt analytics and intelligence-oriented digital technologies. Based on the marginal effects analysis, we found that compared to small-sized health systems, there is a 0.145 decrease in the probability of adopting AODT by medium-sized health systems.

We found a significant and negative relationship between major teaching health systems and AODT ($P<.001$), indicating that compared to major teaching health systems, nonteaching health systems have a greater orientation toward AODT. The marginal effects analysis suggested a 0.123 decrease in the probability of adopting AODT in major teaching health systems than in the nonteaching health systems.

A high-burden hospital also had a significant and negative impact on AODT (*P*<.001). This result indicates that a health system without a high uncompensated care burden hospital is more likely to use analytics technologies. We also examined the marginal effects of this variable. The result indicated a 0.039 decrease in the probability of using AODT by a health system with at least one high uncompensated care burden hospital.

Table 5. Ordered logit model estimation results.[a]

Variables[b]	AODT[c] (pseudo-R^2=0.027)		CODT[d] (pseudo-R^2=0.035)		GODT[e] (pseudo-R^2=0.031)		FEDT[f] (pseudo-R^2=0.048)	
	Coefficient (SE)	*P* value	Coefficient (SE)	*P* value	Coefficient (SE)	*P* value	Coefficient (SE)	*P* value
SIZE_B-MEDIUM	–1.283 (.181)	<.001	–1.510 (.290)	<.001	.005 (.305)	.99	.863 (.495)	.08
SIZE_B-LARGE	–.804 (.407)	.05	–1.058 (.454)	.02	.123 (.390)	.75	.707 (.558)	.21
REGION-MW	.139 (.528)	.79	–.078 (.689)	.91	1.005 (.507)	.05	1.365 (.204)	<.001
REGION-SOUTH	–.009 (.355)	.98	–.386 (.504)	.44	.745 (.370)	.04	1.235 (.174)	<.001
REGION-WEST	.391 (.446)	.38	.604 (.560)	.28	.630 (.474)	.18	2.074 (.817)	.01
TEACHING-MINOR	–.182 (.820)	.82	.145 (.649)	.82	–.072 (.874)	.93	.211 (.920)	.82
TEACHING-MAJOR	–1.004 (.218)	<.001	–.410 (.225)	.07	.268 (1.038)	.80	.879 (.801)	.27
REVENUE-MEDIUM	.341 (.526)	.52	–.487 (.365)	.18	.430 (.877)	.62	–.255 (.061)	<.001
REVENUE-HIGH	–.025 (.258)	.92	–.166 (.084)	.05	–.357 (.124)	.004	–.245 (.062)	<.001
HIGH-DSH-HOSP	.612 (.403)	.13	.220 (.224)	.33	–.045 (.503)	.93	–.891 (.435)	.04
HIGH-BURDEN-SYS	.562 (.697)	.42	.980 (.644)	.13	.200 (.208)	.34	1.018 (.347)	.003
HIGH-BURDEN-HOSP	–.376 (.087)	<.001	–.880 (.250)	<.001	–.016 (.202)	.94	–.784 (.281)	.005
OWNERSHIP	–.280 (1.299)	.83	.153 (2.113)	.94	–4.934 (1.974)	.01	–1.523 (1.356)	.26
PHYSICIANS	–.080 (.054)	.14	–.142 (.104)	.17	.048 (.662)	.94	.342 (.260)	.19
HOSPITALS	.180 (.205)	.38	.288 (.174)	.10	.068 (.118)	.57	–.131 (.136)	.34

[a] The results of the cut-off points are omitted for brevity.

[b] See Table 1 for descriptions of the variable codes.

[c] AODT: analytics-oriented digital technologies.

[d] CODT: customer-oriented digital technologies.

[e] GODT: growth and innovation–oriented digital technologies.

[f] FEDT: futuristic and experimental digital technologies.

We found a significant and negative relationship between the CODT orientation and medium size (*P*<.001) as well as large size (*P*=.05), indicating that smaller-sized health systems are apt to adopt CODT. The marginal effects analysis showed that the probability changes for these two factors were 0.130 and 0.073, respectively.

The significant and negative relationships between major teaching (*P*=.07), high revenue (*P*=.05), and inclusion of a high-burden hospital (*P*<.001) and CODT suggest that nonteaching health systems, low-revenue health systems, and health systems that do not have high-burden hospitals are more likely to adopt digital technologies for their customers. The marginal effects for these three variables were 0.032, 0.013, and 0.072, respectively.

Compared with health systems in the northeast, health systems in the midwest (*P*=.05) and south (*P*=.04) were found to be more likely to adopt GODT. These results reveal the influence of health systems' location on their orientation to GODT. More specifically, marginal effects analysis indicated a 0.006 and 0.005 increase in the probability of adopting GODT for health systems in the midwest and southern states, respectively.

Table 5 also shows negative relationships between GODT and high revenue (*P*=.004) as well as ownership (*P*=.01), suggesting that low-revenue health systems and health systems that are owned by noninvestors tend to use GODT. According to the marginal effects, the more specific tendency changes were 0.003 and 0.477 for these two variables, respectively.

Table 5 also shows that the regions of health systems have significant impacts on the FEDT orientation, with health systems in the midwest (*P*<.001), south (*P*<.001), and west (*P*=.01) being more likely to adopt FEDT than those in the northeast. The changes in the marginal effects were 0.006, 0.006, and 0.007, respectively.

There were positive relationships between FEDT and medium size (*P*=.08) and system-wide burden (*P*=.003), suggesting that medium-sized health systems and health systems that have a high system-wide uncompensated care burden tend to adopt FEDT. The changes in the probability of adopting FEDT were the same for these two variables (0.004).

By contrast, there were negative relationships between FEDT and medium revenue (*P*<.001), high revenue (*P*<.001), the inclusion of hospitals with DSH patients (*P*=.04), and inclusion of high-burden hospitals (*P*=.005). These results indicate that low-revenue health systems, and health systems without a high DSH patient percentage hospital and no high uncompensated care burden hospitals are more inclined to use FEDT. According to the marginal effects analysis, these health systems indicate an increase of 0.001, 0.001, 0.006, and 0.005, respectively, in the probability of adopting FEDT.

Discussion

Principal Findings

This study first explored the digital orientations of health systems across the United States and then examined the factors that may influence the digital orientations of health systems, comparing across the current analytics and customer-oriented technologies, and the growth and futuristic-oriented technologies [27]. The main findings suggest that (1) health systems in the midwest and southern states, along with low-revenue and noninvestor-owned health systems have growth or futuristic digital orientations (ie, GODT or FEDT); and (2) small-sized, nonteaching, and less burdened health systems are still focusing on current digital technologies such as analytics or customer-oriented technologies (ie, AODT or CODT).

The first set of results suggests the impacts of size, teaching status, and burden of a health system on its digital orientations. More specifically, the smaller-sized health systems are more likely to adopt analytics and customer-oriented digital technologies. Plausibly, smaller health systems are constrained by the complexities of digital technologies to explore advanced digital technologies such as artificial intelligence and robotics [28]. Smaller health systems may not have a research and development team or an independent information technology department to steer and improve future technologies and align them to create value.

Second, the findings show that nonteaching health systems tend to focus more on analytics and customer-oriented digital technologies. It is widely known that teaching hospitals are somehow focused on more cutting-edge technologies and experiment on the future technologies compared with nonteaching hospitals that have limited access to such opportunities. Accordingly, the variation in the strategic vision to guide digital orientations is apparent [29,30].

Third, health systems including hospitals with a lower uncompensated care burden are apt to choose AODT or CODT. According to a prior study, uncompensated care decreased at hospitals when there was Medicaid expansion [31]. A low uncompensated care burden health system presumably has higher revenue, and has no strong motivation to be future-thinking. In other words, such health systems are satisfied with the revenue from traditional care avenues through the current basic analytics and customer-oriented digital technologies.

The second set of findings highlights the role of location, revenue, and ownership status of a health system based on its adoption of digital technologies. First, we found that health systems in the midwest and south are more growth and futuristic-oriented. An explanation for this may be that while health systems in the northeast had advanced with respect to records-based digital technologies, systems located in the midwest and south have lagged in this transition [32,33]. While realizing the value potential, systems in the midwest and south may be trying to make up for the lost time to gain a competitive advantage.

Similarly, health systems with a low revenue are also more likely to adopt futuristic and growth-oriented digital technologies. Although it appears counterintuitive, leaders of low-revenue systems have strong motivation to explore and leverage futuristic digital technology to grow rather than risk failure by sustained low revenue. In other words, low-revenue systems are aspiring that the futuristic technologies will help them to be efficient and cost-effective on the digital transformative path.

Third, the results show that noninvestor-owned health systems have higher probabilities of adopting growth and futuristic digital technologies. The reason may be related to one of the following two aspects. On the one hand, it may be that investor-owned hospitals have already spent resources on state-of-the-art digital technologies, and further investment is duplicative. On the other hand, investor-owned systems may be risk-averse given that quarterly earnings are rarely driven by high-cost digital investment [34,35]. Nevertheless, noninvestor-owned health systems find themselves in a unique position to take advantage of investor-owned health systems' slow adoption of future-oriented digital technologies.

Implications

These findings have several practice and policy implications. There are strong indications that small-sized and low-revenue health systems need financial incentives to bridge the digital gap. Although their aspirations are high, current revenues may not allow the investment needed to create a competitive advantage. On a similar note, it is possible that health systems' aspirations will end in unmet expectations unless those expectations are used to guide the health system through effective adoption and utilization of appropriate digital technologies.

Earlier failures of electronic health record implementations by several health systems indicate that digital technology adoption and implementation is a risky venture. Given past failures, evidence suggests that some health systems are either not disciplined enough or financially prepared for such an implementation. We suggest that policymakers pay attention to these past failures and formulate a well-orchestrated, incentive-based approach for these health systems to succeed in the future.

A point of concern here is the observed significant variations of health systems' digital orientations. Although it may be realistic to expect that some variations are unavoidable, such significant variations are of some concern, particularly when all health systems are engaged in similar operations, businesses, and service delivery approaches. This again points to the lack of consensus and specific public policy in the health care sector regarding the development of digital technologies across the health system. It is clear that a top-level US health systems digital strategy and plan, driving all health systems with similar implementation criteria is desperately needed.

Once an appropriate public policy is in place, we believe that the market will drive relevant training opportunities, improve organizational capabilities, and focus the attention of CEOs necessary to drive regional developments. Additionally, we believe that such a process will successfully drive financial, operational, and strategic support for nondigital health systems that cannot thrive in the absence of such support.

At present, it appears that there are only a few senior leaders of health systems who can, without hesitation, state that their system has a health care technology plan and program. Some wonder if they even have a program at all. As these leaders can adapt to a more significant industry-wide policy of digital technology, their systems will be better positioned to move toward the future to help overcome employee-level resistance to these changes.

Limitations and Directions for Future Research

We acknowledge some limitations of this study. First, although we examined the impacts of revenue on digital orientations, we could not capture the actual digital expenditures, which is a more significant factor for health systems' digital options. In the future, we plan to collect data to reflect this factor.

Furthermore, there are several significant barriers to adopting futuristic digital technologies, such as security concerns. Future studies may focus on how these barriers and orientations are aligned.

We also recognize that the underlying tone in this study is that the growth and futuristic orientation is more important than the customer and analytic orientation for health systems, following prior research [4,27]. However, in the current US health care industry, we acknowledge that customer orientation and analytics-driven intelligence also play significant roles in improving quality and efficiency while reducing care delivery costs. Independent of this perspective, a future study may correlate the digital technology orientations with the performance of health systems to justify this assumption. Moreover, we have only focused on the influences of objective factors on digital orientations in this study due to the nature of using secondary data. Future studies may consider other subjective factors such as senior leadership support and strategic alignment–relevant factors.

Conclusions

The challenges and uncertainties that the COVID-19 pandemic presented to health systems in the United States were unprecedented. The pandemic propelled the transformative and disruptive powers of digitalization to the forefront. The unprecedented surge of telehealth with remote and virtual care reshaped delivery models, which changed the relationship between patients and care providers. Further, the pandemic relatively quickly reshaped the acceptance of virtual technology. More than ever before, health care was provided virtually, and patients who used to have to come to a hospital or clinic were free from that burden. Given this change, senior leaders need to understand the digital orientations in their health systems to address the challenges and prepare for the uncertainties.

Almost all health systems have adopted customer-facing digital technologies to enable remote and virtual care deliveries. Indeed, several health systems have analytics-driven decision-making capabilities. Nevertheless, not many health systems use technologies for workflow alignments to spur innovation and futuristic growth. On the one hand, smaller-sized, nonteaching, and low-burdened health systems tend to adopt analytics and customer-oriented digital technologies. The rationale for their choice may be financial constraints, lack of capability, and lack of support with respect to policy or technical support.

Finally, health systems in the midwest and south, along with low-revenue and noninvestor-owned health systems are more likely to adopt futuristic and growth-oriented digital technologies. The underlying reasons can be very complex, but this finding indicates the development pattern regarding location, financial performance, and ownership status. Some traditionally underrepresented health systems are making efforts to grow by leveraging disruptive digital technologies. While this is excellent progress, such efforts need to be supported at the highest echelons of the policy level. With guidance, these policies can better ensure that future failures are avoided.

The response to the disruption of the COVID-19 pandemic highlights the significance of digital technologies. In the post COVID-19 era, we believe that more and more health systems will see the value of digital transformation. However, some health systems may fall back in this process due to resource constraints, including tangible resources such as budget and intangible resources such as information technology capabilities [36]. It is crucial to provide policy and technical assistance to support the future-oriented digital transformation efforts in health systems. We give the clarion call to form a top-level US health systems digital strategy and plan to shape the development blueprints for all health systems and the nation.

Acknowledgments

This research is part of the Health Systems' Climate Study of 2021 conducted by the Health Administration Research Consortium. The Climate Study aims to understand the current state of health systems in the United States following the COVID-19 pandemic. We extend our sincere thanks to the CEOs from 135 health systems for their active and candid participation. The authors thank Naser Shekarian, PhD candidate in the CSIS Business program at the University of Colorado Denver, for help in compiling the addresses of CEOs for the climate study data collection. The authors appreciate the support from the Business School at the University of Colorado Denver for this project, specifically the Health Administrations and the CSIS Business PhD Program for the time and efforts of the authors involved with this study.

Conflicts of Interest

None declared.

References

1. Kruse CS, Beane A. Health information technology continues to show positive effect on medical outcomes: systematic review. J Med Internet Res 2018 Feb 05;20(2):e41 [FREE Full text] [doi: 10.2196/jmir.8793] [Medline: 29402759]
2. Smoller JW. The use of electronic health records for psychiatric phenotyping and genomics. Am J Med Genet B Neuropsychiatr Genet 2018 Oct 30;177(7):601-612 [FREE Full text] [doi: 10.1002/ajmg.b.32548] [Medline: 28557243]
3. Adler-Milstein J, Holmgren AJ, Kralovec P, Worzala C, Searcy T, Patel V. Electronic health record adoption in US hospitals: the emergence of a digital "advanced use" divide. J Am Med Inform Assoc 2017 Nov 01;24(6):1142-1148 [FREE Full text] [doi: 10.1093/jamia/ocx080] [Medline: 29016973]
4. Kindermann B, Beutel S, Garcia de Lomana G, Strese S, Bendig D, Brettel M. Digital orientation: Conceptualization and operationalization of a new strategic orientation. Eur Manag J 2020 Oct:online ahead of print. [doi: 10.1016/j.emj.2020.10.009]
5. Gatignon H, Xuereb J. Strategic orientation of the firm and new product performance. J Market Res 2018 Dec 20;34(1):77-90. [doi: 10.1177/002224379703400107]
6. Newbert SL. Empirical research on the resource-based view of the firm: an assessment and suggestions for future research. Strat Mgmt J 2006 Feb;28(2):121-146. [doi: 10.1002/smj.573]
7. Narver JC, Slater SF. The effect of a market orientation on business profitability. J Market 1990 Oct;54(4):20. [doi: 10.2307/1251757]
8. Covin JG, Slevin DP. Strategic management of small firms in hostile and benign environments. Strat Mgmt J 1989 Jan;10(1):75-87. [doi: 10.1002/smj.4250100107]
9. Schweiger SA, Stettler TR, Baldauf A, Zamudio C. The complementarity of strategic orientations: A meta-analytic synthesis and theory extension. Strat Mgmt J 2019 May 22;40(11):1822-1851. [doi: 10.1002/smj.3042]
10. Kohli R, Melville NP. Digital innovation: a review and synthesis. Info Systems J 2018 Apr 17;29(1):200-223. [doi: 10.1111/isj.12193]
11. Nambisan S, Wright M, Feldman M. The digital transformation of innovation and entrepreneurship: Progress, challenges and key themes. Res Policy 2019 Oct;48(8):103773. [doi: 10.1016/j.respol.2019.03.018]
12. Bharadwaj A, El Sawy OA, Pavlou PA, Venkatraman N. Digital business strategy: toward a next generation of insights. MIS Quart 2013 Feb 2;37(2):471-482. [doi: 10.25300/misq/2013/37:2.3]
13. Drnevich PL, Croson DC. Information technology and business-level strategy: toward an integrated theoretical perspective. MIS Quart 2013 Feb 2;37(2):483-509. [doi: 10.25300/misq/2013/37.2.08]
14. Hsieh A, Yen C, Chin K. Participative customers as partial employees and service provider workload. Int J Serv Ind Manag 2004 Apr 01;15(2):187-199. [doi: 10.1108/09564230410532501]
15. Walker RH, Johnson LW. Why consumers use and do not use technology-enabled services. J Serv Market 2006 Feb;20(2):125-135. [doi: 10.1108/08876040610657057]
16. Choudhury V, Karahanna E. The relative advantage of electronic channels: a multidimensional view. MIS Quart 2008;32(1):179. [doi: 10.2307/25148833]
17. Curran JM, Meuter ML. Self‐service technology adoption: comparing three technologies. J Serv Market 2005 Mar;19(2):103-113. [doi: 10.1108/08876040510591411]
18. Klier J, Klier M, Müller A, Rauch C. The impact of self-service technologies – towards an economic decision model and its application at the German Federal Employment Agency. J Decis Syst 2016 Mar 15;25(2):151-172. [doi: 10.1080/12460125.2016.1141274]
19. Weijters B, Rangarajan D, Falk T, Schillewaert N. Determinants and outcomes of customers' use of self-service technology in a retail setting. J Serv Res 2016 Jun 29;10(1):3-21. [doi: 10.1177/1094670507302990]
20. Bagheri M, Mitchelmore S, Bamiatzi V, Nikolopoulos K. Internationalization orientation in SMEs: the mediating role of technological innovation. J Int Manag 2019 Mar;25(1):121-139. [doi: 10.1016/j.intman.2018.08.002]
21. Chang HH, Wong KH, Chiu WS. The effects of business systems leveraging on supply chain performance: Process innovation and uncertainty as moderators. Inf Manag 2019 Sep;56(6):103140. [doi: 10.1016/j.im.2019.01.002]
22. 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]
23. Cresswell K, Cunningham-Burley S, Sheikh A. Health care robotics: qualitative exploration of key challenges and future directions. J Med Internet Res 2018 Jul 04;20(7):e10410 [FREE Full text] [doi: 10.2196/10410] [Medline: 29973336]
24. van der Schaar M, Alaa AM, Floto A, Gimson A, Scholtes S, Wood A, et al. How artificial intelligence and machine learning can help healthcare systems respond to COVID-19. Mach Learn 2020 Dec 09;110(1):1-14 [FREE Full text] [doi: 10.1007/s10994-020-05928-x] [Medline: 33318723]
25. Health Systems Climate Study Report. CU Denver Business School, University of Colorado Denver. URL: https://business.ucdenver.edu/harc/2021ClimateStudyReport [accessed 2021-08-07]

26. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. J Market Res 2018 Nov 28;18(1):39-50. [doi: 10.1177/002224378101800104]
27. Scuotto V, Arrigo E, Candelo E, Nicotra M. Ambidextrous innovation orientation effected by the digital transformation. Bus Process Mgmt J 2019 Dec 03;26(5):1121-1140. [doi: 10.1108/bpmj-03-2019-0135]
28. Townsend K, McDonald P, Cathcart A. Managing flexible work arrangements in small not-for-profit firms: the influence of organisational size, financial constraints and workforce characteristics. Int J Human Res Manag 2016 Jan 29;28(14):2085-2107. [doi: 10.1080/09585192.2015.1136671]
29. DesRoches CM, Worzala C, Joshi MS, Kralovec PD, Jha AK. Small, nonteaching, and rural hospitals continue to be slow in adopting electronic health record systems. Health Aff (Millwood) 2012 May;31(5):1092-1099. [doi: 10.1377/hlthaff.2012.0153] [Medline: 22535503]
30. Panch T, Szolovits P, Atun R. Artificial intelligence, machine learning and health systems. J Glob Health 2018 Dec;8(2):020303. [doi: 10.7189/jogh.08.020303] [Medline: 30405904]
31. Dranove D, Garthwaite C, Ody C. Uncompensated care decreased at hospitals In Medicaid expansion states but not at hospitals In nonexpansion states. Health Aff (Millwood) 2016 Aug 01;35(8):1471-1479. [doi: 10.1377/hlthaff.2015.1344] [Medline: 27503973]
32. Cronin CE, Bolon DS. Comparing hospital mission statement content in a changing healthcare field. Hosp Top 2018 Sep 20;96(1):28-34. [doi: 10.1080/00185868.2017.1366188] [Medline: 28929942]
33. Goodpasture JE, Rogers N. Standardizing community benefit between not-for-profit and investor-owned hospitals. Int J Healthc Manag 2019 Oct 25;14(2):583-587. [doi: 10.1080/20479700.2019.1679517]
34. Gajendran M, Umapathy C, Loganathan P, Hashash JG, Koutroubakis IE, Binion DG. Analysis of hospital-based emergency department visits for inflammatory bowel disease in the USA. Dig Dis Sci 2016 Feb 30;61(2):389-399. [doi: 10.1007/s10620-015-3895-2] [Medline: 26423080]
35. Vallabhajosyula S, Patlolla SH, Dunlay SM, Prasad A, Bell MR, Jaffe AS, et al. Regional variation in the management and outcomes of acute myocardial infarction with cardiogenic shock in the United States. Circ Heart Failure 2020 Feb;13(2):1-10. [doi: 10.1161/circheartfailure.119.006661]
36. Kamasak R. The contribution of tangible and intangible resources, and capabilities to a firm’s profitability and market performance. Eur J Manag Bus Econ 2017 Jul 17;26(2):252-275. [doi: 10.1108/ejmbe-07-2017-015]

Abbreviations

AODT: analytics-oriented digital technologies
CEO: chief executive officer
CODT: customer-oriented digital technologies
DSH: Disproportionate Share Hospital
FEDT: futuristic and experimental digital technologies
GODT: growth and innovation–oriented digital technologies

Edited by C Basch; submitted 14.05.21; peer-reviewed by T Deng, M Dadgar; comments to author 04.06.21; revised version received 16.06.21; accepted 16.06.21; published 16.08.21.

Please cite as:
Khuntia J, Ning X, Stacey R
Digital Orientation of Health Systems in the Post–COVID-19 “New Normal” in the United States: Cross-sectional Survey
J Med Internet Res 2021;23(8):e30453
URL: https://www.jmir.org/2021/8/e30453
doi:10.2196/30453
PMID:34254947

©Jiban Khuntia, Xue Ning, Rulon Stacey. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 16.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Characterization of Global Research Trends and Prospects on Single-Cell Sequencing Technology: Bibliometric Analysis

Quan Wang[1,2*], MD, PhD; Ke-Lu Yang[3*], MN; Zhen Zhang[1,2], PhD; Zhu Wang[4], MD, PhD; Chen Li[1,2], BA; Lun Li[5], MD, PhD; Jin-Hui Tian[6], PhD; Ying-Jiang Ye[1,2], MD, PhD; Shan Wang[1,2], MD, PhD; Ke-Wei Jiang[1,2*], MD

[1]Department of Gastroenterological Surgery, Peking University People's Hospital, Beijing, China

[2]Laboratory of Surgical Oncology, Beijing Key Laboratory of Colorectal Cancer Diagnosis and Treatment Research, Peking University People's Hospital, Beijing, China

[3]Evidence-Based Nursing Center, School of Nursing, Lanzhou University, Lanzhou, China

[4]Department of Gastrointestinal Surgery, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, China

[5]Department of Breast Surgery, Shanghai Cancer Center, Fudan University, Shanghai, China

[6]Evidence-Based Medicine Center, School of Basic Medical Sciences, Lanzhou University, Lanzhou, China

[*]these authors contributed equally

Corresponding Author:
Ke-Wei Jiang, MD
Department of Gastroenterological Surgery
Peking University People's Hospital
No.11 Xizhimen South Street, Xicheng District
Beijing, 100044
China
Phone: 86 010 88326600
Email: jiangkewei@pkuph.edu.cn

Abstract

Background: As single-cell sequencing technology has been gradually introduced, it is essential to characterize global collaboration networks and map development trends over the past 20 years.

Objective: The aim of this paper was to illustrate collaboration in the field of single-cell sequencing methods and explore key topics and future directions.

Methods: Bibliometric analyses were conducted with CiteSpace and VOSviewer software on publications prior to November 2019 from the Web of Science Core Collection about single-cell sequencing methods.

Results: Ultimately, we identified 2489 records, which were published in 495 journals by 14,202 authors from 1970 institutes in 61 countries. There was a noticeable increase in publications in 2014. The United States and high-income countries in Europe contributed to most of the records included. Harvard University, Stanford University, Karolinska Institutes, Peking University, and the University of Washington were the biggest nodes in every cluster of the collaboration network, and SA Teichmann, JC Marioni, A Regev, and FC Tang were the top-producing authors. Keywords co-occurrence analysis suggested applications in immunology as a developing research trend.

Conclusions: We concluded that the global collaboration network was unformed and that high-income countries contributed more to the rapidly growth of publications of single-cell sequencing technology. Furthermore, the application in immunology might be the next research hotspot and developmental direction.

(J Med Internet Res 2021;23(8):e25789) doi:10.2196/25789

KEYWORDS

single-cell sequencing; bibliometric analysis; cancer; cancer genomics; bioinformatics; cancer subtyping; tumor dissociation; tumor microenvironment; precision medicine; immunology; development trends; hotspots; research topics; Web of Science; CiteSpace; VOSviewer; network

Introduction

A single cell is regarded as the fundamental unit of an organism. Moreover, there are approximately 10^{14} single cells that comprise the complex tissues and organs of *Homo sapiens* [1]. In order to study genomic or transcriptome information, most published studies to date have collected data by analyzing bulk tissue samples and true genomic information representing the average of millions of cells [2]. Accordingly, because of cell-to-cell variability, traditional sequencing methods only provide the average diversity of cells, instead of obtaining information on entire cellular heterogeneity [3]. When facing the complexity of disease, it is difficult for these averaged data sets to represent cell-to-cell variations; therefore, it is difficult for researchers to identify rare cells, including cancer stem cells, that play a key role in cancer progression. To avoid the weakness and limitations of bulk sequencing technologies, the single-cell method offers a novel possibility that focuses on the single-cell level [4], with detailed and comprehensive studies of individual cells rather than traditional analysis of bulk tissue. The first step of sequencing a single cell, however, involves capturing individual cells; therefore, it is of utmost importance to establish various methods for isolating single cells from abundant populations or rare single cells (<1%) from typical populations. Several approaches have been well-established, including mouth pipetting, serial dilution, robotic micromanipulation, flow-assisted cell sorting, microfluidic platforms, and laser-capture microdissection, however, the isolation of rare single cells (<1%) is far more challenging [2].

With increasing developments of high-throughput sequencing technologies over the past couple of decades, an increasing number of commercial platforms have been introduced for experimental and clinical applications, and thousands of genomes from numerous species have been sequenced [3]. Researchers now have a wealth of data to explore genomic or transcriptome information. Single-cell sequencing refers to the sequencing of a single-cell genome or transcriptome, to reveal cell population differences and cellular evolutionary changes [5]. Furthermore, these commercial methods interrogate the single-cell genome, transcriptome, and epigenome. For example, since the introduction of the first single-cell RNA-sequencing technique in 2009 [6], single-cell RNA sequencing has become a powerful and useful approach to study individual cell transcriptomes on a large scale [7], similar to the technology of multiomics sequencing [8]. When it comes to the applications of single-cell sequencing techniques, the important role of single-cell sequencing in various fields, including oncology [9], microbiology [10], neurology [11,12], reproduction [13], and immunology [14] has been highlighted. These rapidly developing methods have and will continue to lead to new discoveries in many scientific fields.

Recently, because single-cell sequencing technology has been gradually introduced for clinical applications, there has been an increase in research papers published online on this topic. There are more than 70 related records in this field, published online each month on PubMed, and these are of a variety of publication types. Some papers offered a brief introduction of single-cell sequencing technology [15], whereas others focused on investigating its applications [16]. In some reviews [7,17], the current strengths and disadvantages of single-cell sequencing technology were summarized. Published original articles updated or simplified the algorithms of computational methods [18].

Based on this, it is important for us to better understand and learn more from the developmental trends and novel advancements of single-cell sequencing techniques macroscopically; however, rapidly learning about the current landscape of specific biomedical research is still a challenge. Fortunately, the emergence of bibliometric analysis provides an approach to statistically and quantitatively visualize evidence according to the information in published records in a specific field [19,20], which includes data, such as keywords, citation reports, authors, affiliated countries, institutions, and journals, from a large number of published studies. To the best of our knowledge, no specific study has focused on the characterization of research hotspots, global research collaborations, and developmental trends of single-cell sequencing techniques. To characterize current evidence and establish future research directions, it is essential to perform bibliometric analysis to map global collaboration patterns and developmental trends of published single-cell sequencing methods.

The aims of this bibliometric analysis were to map this research landscape through analysis of the information in published records.

Methods

Data Source and Search Strategy

We searched the Web of Science Core Collection (WoSCC) without publication date restrictions on November 12, 2019. We updated the database search on January 29, 2020, to complete data retrieval for the year of 2019; however, only the number of records published annually was updated.

The search was performed using the following keywords and terms: (“single-cell RNA sequencing” OR “scRNA-seq” OR “single-cell RNA-seq” OR “single-cell sequencing” OR “single-cell transcriptomic” OR “single-cell ATAC” OR “single-cell RNA-sequencing” OR “single-cell omics sequencing” OR “scRNA seq”). The search strategy was peer-reviewed and guided by TJH (who has over 10 years of experience as an information specialist). In this study, only publications in English were included; however, there was no restriction on data category.

Inclusion and Exclusion Criteria

Inclusion criteria were published records, including articles, comments, reviews, letters, and brief introductions on single-cell sequencing techniques. We included the records that, not only focused on the single-cell sequencing technology after the paper of single-cell RNA-sequencing published in 2009 [6], but also, generally used the single-cell sorting methods such as Flow sorting in early time. We excluded duplicates, conference abstracts, and manuscripts in a language other than English, for consistent and accurate information collection from multianalysis results based on published records. Two groups of reviewers (WQ/WZ/JKW and YKL/LL) independently screened the titles and abstracts to select the articles after standard selection training. Full-texts were retrieved when necessary. Disagreements were discussed and solved between the 2 groups.

Social Network Map Software

CiteSpace (version 5.3.R4) was used for social network analysis of developmental dynamics, future trends, hotspots, and key points in scientific literature of a specific topic [21]. Burst detection, to identify keywords and references that appear with an abrupt change in frequency at a certain period, were considered to be hot keywords or references at that point in time. Clustering (or co-citation, ie, both A and B are cited by C) analysis of references was also conducted [21]. We used VOSviewer (version 1.6.9; Leiden University) to visualize the collaborations between authors of a list of publications as well as those between countries, institutions, and high frequency keywords [22]. One classification method uses a metric based on the number of co-authored articles, which allowed authors, institutions, countries or keywords to be clustered, where those belonging to the same group cooperated more with each other [23]. In network maps, unequal-size nodes with different colors represent differences in the number or frequency of published records in clusters among the same research topics [21]. Lines between nodes indicate the strength of collaborations (the stronger the collaboration, the thicker the line) [24]. We also used overlay visualization, in which the color of the node represents the average year in which each author, institution, country or keyword was used [25].

Data were saved as *Plain Text* with *Full Record and Cited References* in WoSCC and imported into CiteSpace.

Statistical Analysis

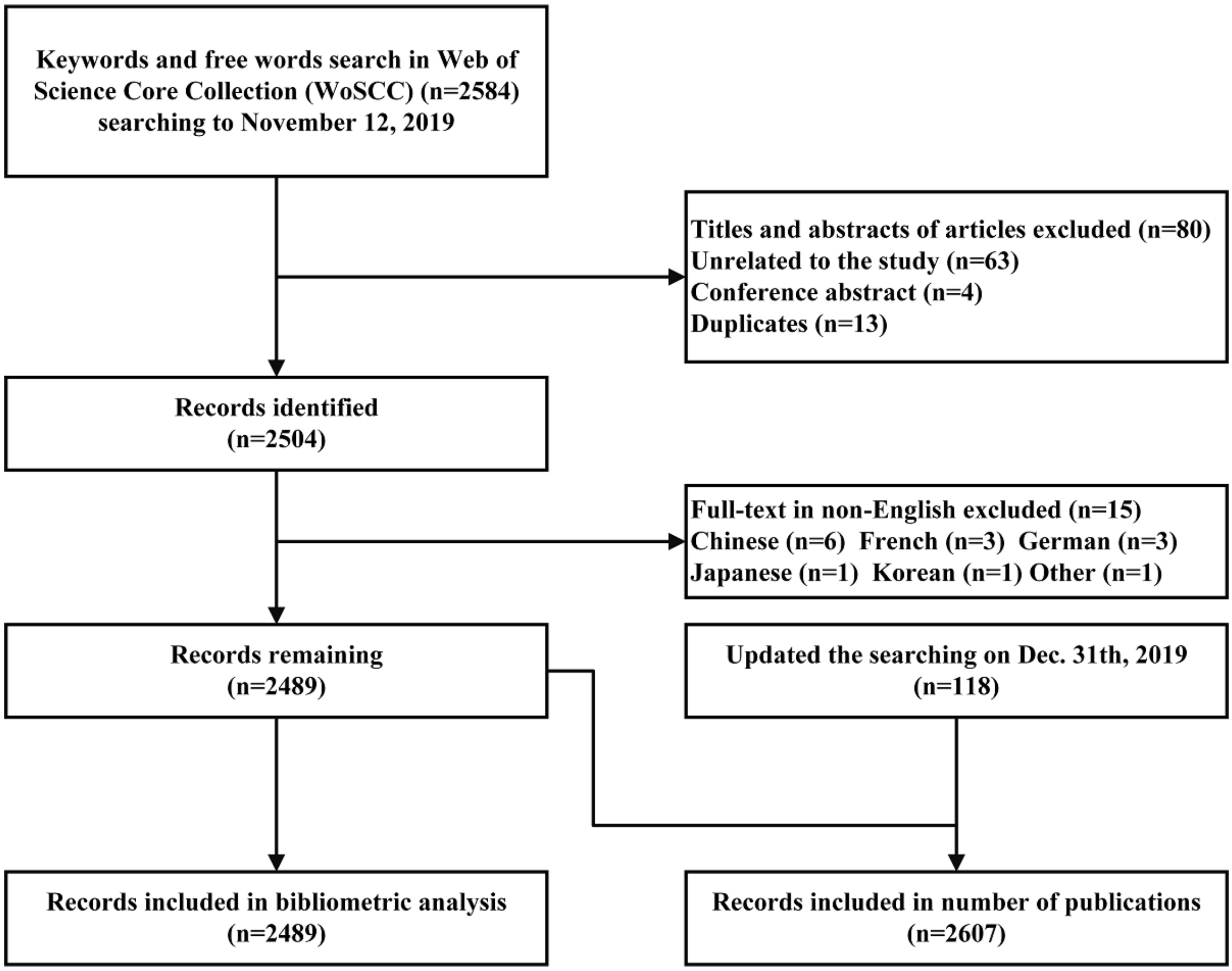
SPSS software (version 22.0; IBM Corp) and Origin (version 2018; OriginLab Corp) were used for data analysis. Continuous variables were presented as mean and standard deviation or median and interquartile range, and categorical variables were expressed using frequencies and percentages.

Results

Search Results

A total of 2584 publications were identified from WoSCC. After screening titles and abstracts, duplicates, unrelated topics, and conference abstracts (n=80) or non-English publications (n=15) were excluded. Finally, 2489 studies were included for bibliometric analysis (Figure 1); 2607 was identified as the number of records published.

Figure 1. Record identification and selection.


Keywords and free words search in Web of Science Core Collection (WoSCC) (n=2584) searching to November 12, 2019
Titles and abstracts of articles excluded (n=80)
Unrelated to the study (n=63)
Conference abstract (n=4)
Duplicates (n=13)
Records identified (n=2504)
Full-text in non-English excluded (n=15)
Chinese (n=6) French (n=3) German (n=3)
Japanese (n=1) Korean (n=1) Other (n=1)
Records remaining (n=2489)
Updated the searching on Dec. 31th, 2019 (n=118)
Records included in bibliometric analysis (n=2489)
Records included in number of publications (n=2607)

Global Publication Trends

Papers published over the past of two decades (Table 1 and Table 2) were primarily classified into 7 types. Most papers were original articles (1858/2607, 71.27%), and there were 388/2607 meeting abstracts (14.84%) and 217/2607 reviews (8.32%). Most publication types were articles, meeting abstracts, and reviews, which can greatly reflect the development trends and changes in single-cell sequencing technology. The first publication on single-cell sequencing technology was in 2001 [26], which reported single-cell sequencing of dinoflagellate (Class Dinophyceae) nuclear ribosomal genes. In addition, there were only was only 1 to 3 publication per year before 2010. From 2010 to 2019, the number of publications showed a noticeable upward trend, with increases since the year 2014 (n=59), reaching 987 publications in 2019.

Table 1. Number of published records per year.

Year	Records, n
2001	1
2002	0
2003	2
2004	0
2005	1
2006	1
2007	1
2008	2
2009	2
2010	3
2011	10
2012	21
2013	35
2014	59
2015	127
2016	225
2017	406
2018	722
2019	989
Total	2607

Table 2. Percentage of published records by publication type.

Publication type	Records (n=2607), n (%)
Article	1858 (71.27)
Meeting abstract	388 (14.84)
Review	217 (8.32)
Editorial material	97 (3.72)
Correction	29 (1.11)
Letter	13 (0.54)
News item	5 (0.19)

Countries and Districts

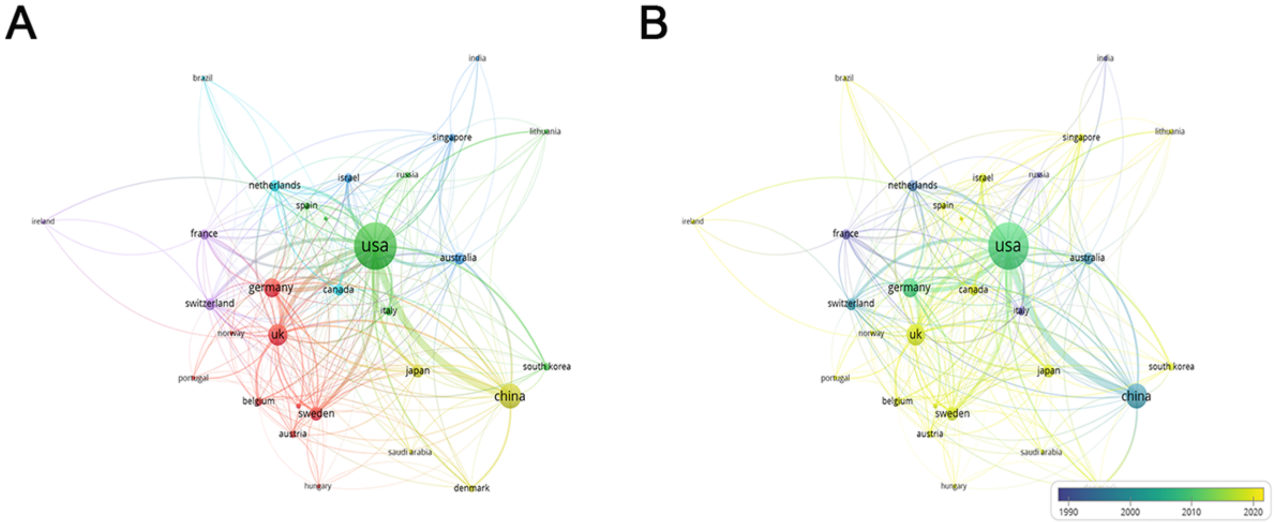
In total, 61 countries contributed to publications on single-cell sequencing technology worldwide. The results demonstrated that 7 countries published more than 100 records (Table 3), and the United States (1454/2489, 58.4%), with more than half of all identified records, published the most. The top 2 to 7 were China (414/2489, 16.6%), the United Kingdom (305/2489, 12.3%), Germany (225/2489, 9.0%), Sweden (134/2489, 5.4%), Japan (108/2489, 4.3%), and Switzerland (104/2489, 4.2%). Although the number of documents published by Australia, Canada, and the Netherlands was less than 100, no significant differences were observed (*P*=.10), and the contribution was almost equal to that of Japan and Switzerland. In addition, the cooperation among studies in China was a little less than that in the United Kingdom. However, Sweden had the highest average number of citations (52.0 times), followed by Canada (37.4 times), the United Kingdom (29.6 times), the United States (27.7 times), Australia (24.7 times), and Germany (23.9 times).

Table 3. Distribution by country.

Rank	Country	Documents, n (%)	Citations, n	Average citations	Total link strength
1	USA	1454 (58.4)	40337	27.7	549
2	China	414 (16.6)	6537	15.8	206
3	UK	305 (12.3)	9026	29.6	214
4	Germany	225 (9.0)	5384	23.9	162
5	Sweden	134 (5.4)	6973	52.0	88
6	Japan	108 (4.3)	1454	13.5	68
7	Switzerland	104 (4.2)	1802	17.3	72
8	Australia	99 (4.0)	2442	24.7	70
9	Canada	85 (3.4)	3177	37.4	53
10	Netherlands	85 (3.4)	1330	15.7	67

For the top 30 most prolific countries, which formed 6 clusters, a network map was created. There were active collaborations among these countries, especially between the United States and China (Figure 2A). From the dynamics and trends, it can be seen that the United States, China, Germany, Switzerland, France, Australia, and the Netherlands have carried out studies on single-cell sequencing technology dating back to 2009. After 2015, many other investigators worldwide started to pay more attention to this field (Figure 2B).

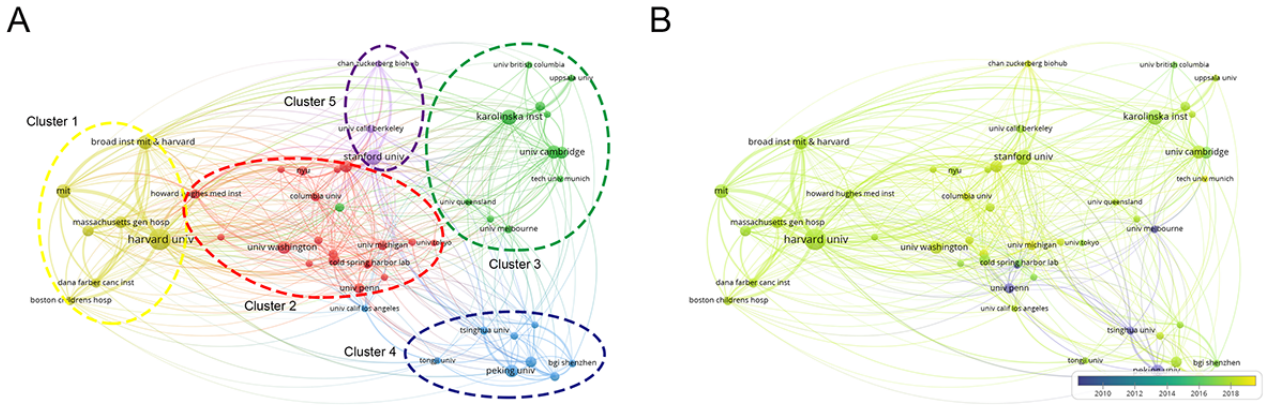
Figure 2. Distribution of countries and regions. (A) The network map of countries (TOP30). (B) Dynamics and trends of countries/regions over years (TOP30).



Universities and Institutions

A total of 1970 universities or institutions made contributions to single-cell sequencing technology; extensive cooperation network analysis was carried out among universities or institutions. In a network map and overlay visualization of institutions with the top 50 frequency, 5 clusters were formed (Figure 3A), and Harvard University (n=226, 9.1%, Cluster 1), Stanford University (n=114, 4.6%, Cluster 5), Karolinska Institute (n=111, 4.5%, Cluster 3), Peking University (n=81, 3.3%, Cluster 4), and the University of Washington (n=77, 3.1%, Cluster 2) were the biggest nodes in every cluster, respectively. Cluster 1 was the biggest cluster, which contained 19 nodes that represented its related different universities or institutions, while Cluster 5 was the smallest, which included 3 nodes. The other clusters included 7 nodes (Cluster 2), 12 nodes (Cluster 3), and 9 nodes (Cluster 4). Furthermore, the results also showed that 4 universities contributed to primary and basic research on these topics earlier in time and started earlier on this field before 2010, which included Peking University, University of Melbourne, University of Penn, and the Tsinghua University. Since then, these studies have gradually gained popularity among other universities and institutes (Figure 3B).

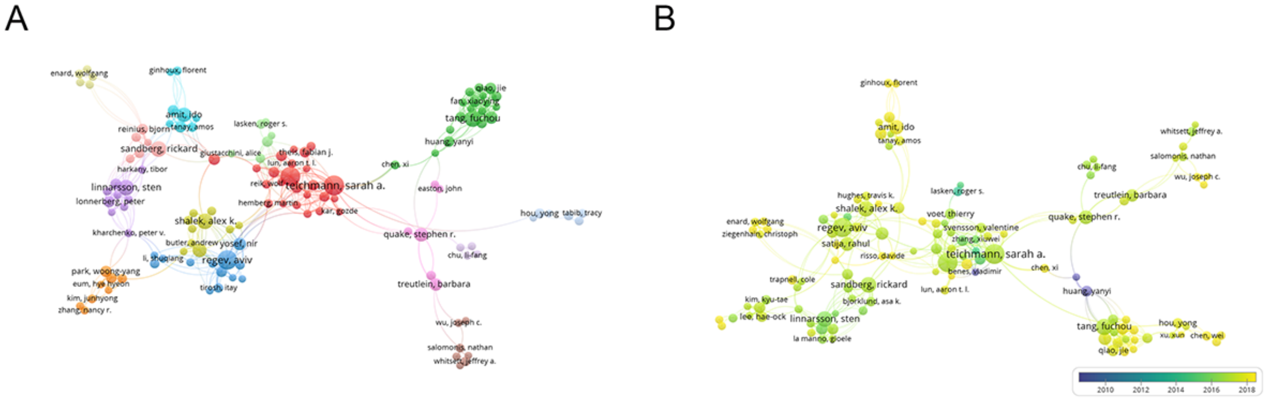
Figure 3. Distribution of universities and institutions. (A) The network map of universities and institutions (TOP50). (B) Dynamics and trends of universities and institutions over time (TOP50).



Authors

A total of 14,202 authors contributed to single-cell sequencing technology publications. The network map and overlay visualization of the top 200 cooperatively productive authors formed several clusters (Figure 4). Of these authors, 5 scientific teams contributed to most publications worldwide. The biggest 3 nodes were SA Teichmann (40/14,202, 1.6%) and JC Marioni 38/14,202, 1.5%), who had the biggest cluster and were from the University of Cambridge. In addition, A Regev (38/14,202, 1.5%) had the second biggest cluster and was from Massachusetts Institute of Technology. The fourth node, FC Tang, represented a group from Peking University (26/14,202, 1.0%), and 2 authors, R Sandberg and S Linnarsson from Sweden contributed to 49 records in total. Moreover, I Amit, from Israel, published 21 records to date (Multimedia Appendix 1).

Figure 4. Distribution of authors. (A) The network map of authors (TOP200). (B) Dynamics and trends of authors over time (TOP200).



Journals

This study showed that a total of 495 journals published articles about single-cell sequencing technology. The results showed that the most productive journal was *Nature Communications* (127/14,202, 5.1%), the most cited journal was *Nature Methods* (55/14,202, 2.2%, 1764 citation times in total), and the journal with the highest average citation was *Nature Biotechnology* (28/14,202, 1.1%, average citation was 56.5 times). The top productive journals, top cited journals, and journals of top average citations are listed (n=20) in Multimedia Appendix 2.

Top 10 Citations of Included Records

Of the 2489 papers, the top 10 publications ranked by citation are listed in Table 4. The first paper was published in the *Journal of Computational Biology* by A Bankevich et al in 2012 [27] and reported a new genome algorithm and its applications to single-cell sequencing, with 5668 citations, which was much higher than that of the second paper (1307 citations) [28]. Of the 10 records, 4 were published in *Science*, and the topics included intratumoral heterogeneity in primary glioblastoma [29], metastatic melanoma [30], and cell types in the mouse cortex and hippocampus revealed by single-cell RNA-sequencing [31]. The latest paper [30], which was published in 2016, has been cited 652 times (from 2016 to 2019).

Table 4. Distribution by top 10-record co-citations.

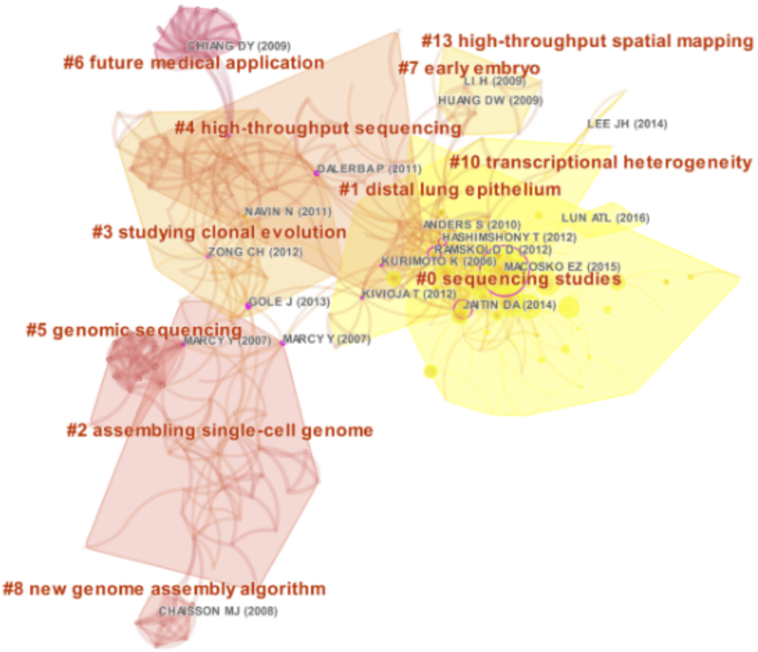
Rank	Top-cited record	Number of citations	Title	Journal
1	A Bankevich, 2012 [27]	5668	SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing	*Journal of Computational Biology*
2	N Navin, 2011 [28]	1307	Tumour evolution inferred by single-cell sequencing	*Nature*
3	AP Patel, 2014 [29]	1168	Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma	*Science*
4	Y Peng, 2012	1039	IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth	*Bioinformatics*
5	A Zeisel, 2015 [31]	923	Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq	*Science*
6	C Trapnell, 2014	773	The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells	*Nature Biotechnology*
7	S Picelli, 2014	734	Full-length RNA-seq from single cells using Smart-seq2	*Nature Protocols*
8	I Tirosh, 2016 [30]	652	Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq	*Science*
9	DA Jaitin, 2014	615	Massively Parallel Single-Cell RNA-Seq for Marker-Free Decomposition of Tissues into Cell Types	*Science*
10	R Satija, 2015	605	Spatial reconstruction of single-cell gene expression data	*Nature Biotechnology*

Co-citation References

We used CiteSpace to visualize the co-citation network of references, which were divided into 11 co-citation clusters (Figure 5). The clusters were listed from 2008 to 2016: "transcriptional heterogeneity" (Cluster 10, n=4), "sequencing studies" (Cluster 0, n=53), "high-throughput spatial mapping" (Cluster 13, n=3), "distal lung epithelium" (Cluster 1, n=37), "studying clonal evolution" (Cluster 3, n=33), "early embryo" (Cluster 7, n=9), "assembling single-cell genome" (Cluster 2, n=34), "high-throughput sequencing" (Cluster 4, n=27), "new genome assembly algorithm" (Cluster 8, n=7), "genomic sequencing" (Cluster 5, n=14), and "future medical applications" (Cluster 6, n=13).

The top 36 references with the strongest citation bursts are presented in Multimedia Appendix 3. The first reference with a citation burst appeared in 2011, and most of the bursts appeared between 2011 and 2016. Only 1 reference [32] appears with a burst in the last 3 years.

Figure 5. The co-citation network of references.


CHIANG DY (2009)
#6 future medical application
#13 high-throughput spatial mapping
#7 early embryo
LI H (2009)
HUANG DW (2009)
#4 high-throughput sequencing
LEE JH (2014)
DALERBA P (2011)
#10 transcriptional heterogeneity
#1 distal lung epithelium
NAVIN N (2011)
LUN ATL (2016)
ANDERS S (2010)
#3 studying clonal evolution
HASHIMSHONY T (2012)
RAMSKOLD D (2012)
ZONG CH (2012)
KURIMOTO K (2006)
MACOSKO EZ (2015)
#0 sequencing studies
KIVIOJA T (2012)
GOLE J (2013)
JAITIN DA (2014)
#5 genomic sequencing
MARCY Y (2007)
MARCY Y (2007)
#2 assembling single-cell genome
#8 new genome assembly algorithm
CHAISSON MJ (2008)

Analysis of Keywords

From the 2489 published records, a total of 6012 keywords were extracted. The network map of the top 100 frequency keywords was clustered and formed 4 clusters (Figure 6A), and the biggest 3 nodes were "gene-expression," "RNA-seq," and "heterogeneity." Moreover, the overlay visualization of the top 100 frequency keywords between 2001 and 2020 is shown in Figure 6B.

Cluster 1, with the biggest node "gene-expression," represented the primary research points of single-cell sequencing, such as stem cells, self-renewal, progenitors, lineage, differentiation, protein and identification, which might be based upon next generation sequencing and single cell isolation methods. Cluster 2 represented improvements in computer algorithms for sequencing data and advancements in single-cell sequencing methods, for instance, keywords such as heterogeneity, transcriptome, RNA-sequencing data, dynamics, noise, and normalization, especially, the stem cells of embryos might be paid more attention to in this cluster. Cluster 3, considered as the currently being developed field in this topic, and could be named the research hotspots of single-cell sequencing. Moreover, the biggest node of Cluster 3 was marked as cancer, which was related to many keywords, including circulating tumor-cells, origin, evolution, landscape, metastasis, resistance, and progression. In addition, it was also indicated that these topics on breast cancer might stay in a leading position and published more papers. Regarding the future direction of single-cell sequencing technology, Cluster 4 was entitled developmental trends of immunology in single-cell sequencing, and keywords included dendritic cells, T-cells, macrophages, and diversity, which might be related to tumor survival, neuronal growth, and activation of regulators. Moreover, the results in Figure 6B showed that the biggest nodes, such as "gene-expression," "RNA-seq," and "heterogeneity," emerged around the year 2005. Subsequently, the frequency of many keywords, including "single-cell," "noise," "seq," and "gene" gradually increased after 2010 and rapidly gained interest from 2015 to 2020.

The data showed that all keywords with citation bursts first appeared in 2001 (Multimedia Appendix 4). Similarly, breast cancer was a primary research hotspot from 2011 to 2017. Furthermore, at the beginning of 2013, many keywords related to computer algorithms, such as "noise" and "evolution," displayed the strongest citation bursts. Regarding the keyword single-cell sequencing, the time-period of the strongest citation bursts was from 2015 to 2017.

Figure 6. Distribution of 100 frequency keywords. (A) Network map of the frequency keywords(TOP100). (B) Dynamics and trends of the frequency keywords over times(TOP100).

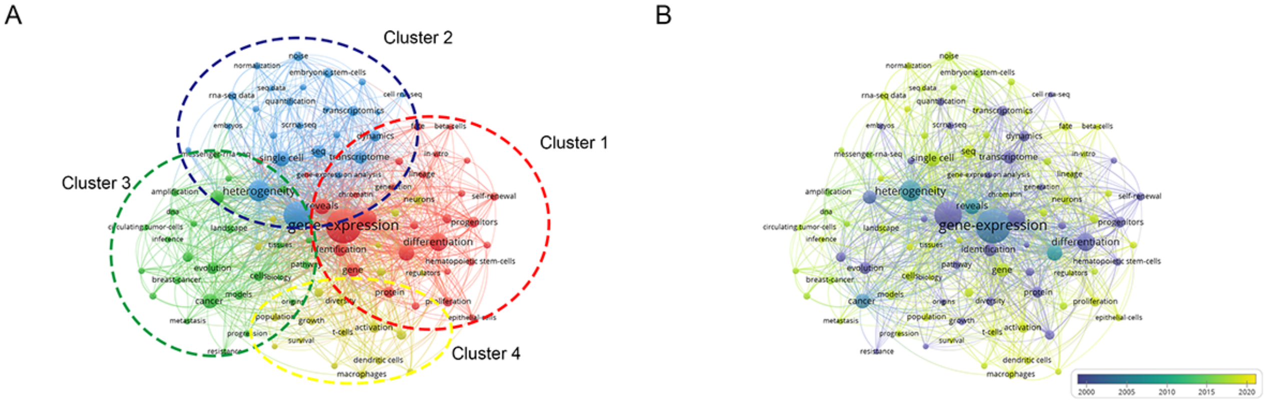


Table 5 summarizes occurrence and the total link strength of the top 30 keywords. Higher keyword occurrences indicated the primary research hotspot was single-cell sequencing applications, including "heterogeneity," "differentiation," "genome," and "transcriptome" of "single cells," which mainly covers the field of stem cells, such as embryonic stem cells, and the field of cancer. Furthermore, the dynamics, landscape, and diversity of single cells maintained their popularity and showed a stronger total link strength among these 30 keywords.

Table 5. Occurrences and total link strength of top 30 keywords.

Rank	Keyword	Occurrences	Total link strength
1	Gene-expression	857	851
2	RNA-seq	563	559
3	Heterogeneity	331	325
4	Stem-cells	235	234
5	Differentiation	231	231
6	Single cell	177	177
7	Reveals	170	170
8	Cancer	155	154
9	Transcriptome	135	135
10	Gene	133	132
11	Mouse	133	132
12	Genome	132	130
13	Identification	125	125
14	Seq	125	122
15	Evolution	115	115
16	Progenitors	93	93
17	Protein	92	92
18	Dynamics	90	90
19	Activation	83	83
20	Transcriptomics	80	78
21	Cells	77	77
22	Receptors	73	73
23	Noise	70	70
24	Mutations	69	69
25	Proliferation	67	67
26	Embryonic Stem-cells	66	66
27	Landscape	65	65
28	Neurons	65	65
29	Models	64	64
30	Diversity	63	63

Discussion

Principal Results

In this study, we summarized research collaborations, new development, and research trends of single-cell sequencing worldwide. In particular, bibliometric analysis offers a new possibility of visualizing current hotpots and trends in this field based on the information of published records [19,20]. We were able to analyze the most productive authors and institutions, by journal and by co-citation references, and we illustrated the different research groups of authors with collaboration networks in the past two decades. In total, 2489 records were identified, published in 495 journals by 14,202 authors from 1970 institutes in 61 countries. The United States had the highest absolute productivity ranking, followed by China and Japan (East Asia), the United Kingdom, Germany, other European countries (Sweden, Switzerland, the Netherlands), and Australia and Canada. Taken together, these results indicated that single-cell sequencing technology developed rapidly and spread globally in the past 5 years (after 2015), which might be influenced by the amount of funding available for single-cell sequencing projects in each country. Li et al [33] investigated the distribution and development of single-cell sequencing in China, the United Kingdom, and the United States, focusing on government support, and concluded that there was a gap between high-income and low-income countries with respect to the amount of funding and number of projects. When compared with China, the amount of funding and the number of projects for single-cell sequencing in the United States and the United

Kingdom had increased dramatically before 2016 (915 projects, cumulative funding of US $539 million). On the other hand, the later the starting time of foundation support for studies on single-cell sequencing, the less applications of this technology.

After the first record on single-cell sequencing, on the nuclear ribosomal genes of dinoflagellate (Dinophyceae), was published in 2001 [26], the number of publications was sporadic (1 to 3 annually until 2010). With the advancement of high-throughput sequencing and new computer algorithms, the development trend of publications changed from 2010 to 2019, showing a clear upward trend since 2014. However, as the development of single-cell sequencing technology has lasted for almost 20 years, several studies represent basic milestones of single-cell sequencing [2]. For example, the advancement of next-generation sequencing platforms in 2005 [34] offered a better approach for primary single-cell sequencing for isolated single cells. Subsequently, Tang et al [6] were the first to introduce RNA transcriptome sequencing of a mammalian cell in 2009, and Navin et al [28] published the first genomic sequencing of single human cells in 2011. Furthermore, several representative papers were published in the subsequent 3 years, including not only exome single-cell DNA sequencing [35] and its related single-cell DNA sequencing of sperm cells [36], neurons [37], circulating tumor cells [38], and the microbial tree [39], but also, the first single-cell sequencing of the RNA of immune cells [40] and of epigenomes [41]. Certainly, single-cell sequencing methods have supported advances in other technologies, including multiomics sequencing [8]; methods have led to novel discoveries in many scientific fields, such as tumor research [9], microbiology [10], neurology [11,12], reproduction [13], and immunology [14], which could explain the recent exponential increase in the number of publications.

While the United States published the most records worldwide (1454/2489, 58.42%), Sweden had the highest citation average (52.04 times). Citations differed from the number of documents, as did the cooperation of countries and regions: United States, China, United Kingdom, and Germany showed the strongest cooperation, with higher total link strengths. This was mainly because studies on single-cell sequencing technology were published earlier in these countries, which was indicated by the dynamics and trends of countries and regions over the years (Figure 3B). We believe that publications from Sweden had deeper impacts on the single-cell sequencing technology, although the field seems to be small and young enough that most of the findings can be entirely due to individual pioneering researchers. Furthermore, we also found that Karolinska Institutes, from Sweden, was one of the biggest nodes in its cluster of collaboration network (Figure 4A; Cluster 3), which might provide an indirect explanation. Similarly, the analysis of institutions and authors almost matched the distribution by countries and districts. Of the 1970 universities or institutions, the institutes from United States (Harvard University, Stanford University, University of Washington, and University of Penn), European universities (Karolinska Institute and Cambridge University), and Chinese institutes (Peking University and Tsinghua University), as well as the University of Melbourne in Australia, have taken leading places in single-cell sequencing technology, with steady collaborations in global groups. Moreover, of the 14,202 authors who contributed to publications, the top 200 authors were situated in the collaboration networks that account for large proportions of research, mainly from the above-mentioned countries and institutes.

Single-cell sequencing avoids the drawbacks of bulk experiments, which only demonstrate average gene expression, and it can be used to study real differences and evolutionary changes in single cells [7]; therefore, it has gained the popularity and interest of the editors of many journals. Of the 495 journals, 20 journals contributed the most. The top journals were *Nature Communications* (most papers with 127 papers), *Nature Methods* (most cited with 1764 citations), and *Nature Biotechnology* (highest average citations with 56.46 citations per paper). Moreover, of included records, all the top 10 most cited papers were published in the top journals. In addition, the research topics of these top-cited records reported or introduced a new genome algorithm and application of single-cell sequencing [27-31], which together have been cited more than 600 times in the past of 5 years. In co-citation analysis, knowledge on “clonal evolution,” “early embryo,” “assembling single-cell genome,” “high-throughput sequencing,” and “new genome assembly algorithm” were the basis of single-cell sequencing research, and most references with citation bursts appeared between 2011 and 2016.

Of 4 clusters in visualization maps to explore key topics and future directions, Cluster 1 represented the basis of single-cell sequencing (next generation sequencing and single cell isolation methods researched earlier), while Cluster 2 represented the development of computer methods and algorithms. Cluster 3 showed current research hotspots of single-cell sequencing, such as applications in cancer research, with keywords including, circulating tumor-cells, origin, evolution, landscape, metastasis, resistance, and progression, and Cluster 4 showed that immunology will be the next hotspot of single-cell sequencing technology research, with keywords including dendritic cells, t-cells, macrophages, and diversity. In particular, in immunology, neuronal growth and the activation of regulators have been studied in recent years, rapidly gaining interest from 2015 to 2020.

Limitations

Our study had several limitations. First, we only searched records from the WoSCC and only included records in English, which might result in selection bias. However, the number of included records in our study was large enough to represent the current evidence landscape of single-cell sequencing technology [42]. Second, although the data had been manually standardized, bias might still exist due to the authors with the same name or the keywords of various expressions. Third, by using software, we might have overlooked some information, which may result in errors in data analysis, even if there are some reasonable and unavoidable differences on the same outcomes between CiteSpace and VOSviewer. This was the first study to perform bibliometric analysis on single-cell sequencing technology research, to identify collaboration networks among authors, countries, and institutions and illustrate developmental trends, current hotspots, and future directions in this field.

Recommendations for Future Work

Single-cell sequencing methods have addressed the drawbacks of averaged gene expression in bulk experiments and provide a possible approach to understanding biological diversity and rare cells. Over the past 10 years, the applications have had a broad impact in many fields, which emerged as delineating cell diversity, tracing cell lineages, classifying cell types, and profiling rare cells [2]. We only selected publications related to immunotherapy in the entire year of 2019, to form and label clusters (Multimedia Appendix 5). Eventually, the formed clusters illustrated that current hotspots in immunology involve at least 5 topics. Cancer immunotherapy is an exciting topic, for which single-cell sequencing methods have great potential for investigating the heterogeneity of intratumor immunity and other topics such as immune cell involvement of breast cancer therapy, role of immune cells in nervous system, formation and differentiation of immune cells, and functions of immune cells in the tumor microenvironment. Cancer therapy research to clarifying the functions of T-cells and macrophages, understand the antigenicity in intratumor heterogeneity, and study the role of clonal diversity in transformation, invasion, and evolution of chemoresistance would open new avenues for preventing cancer in the future. Cancer immunotherapy has revolutionized cancer treatment, and a detailed immune cell atlas at single-cell resolution could facilitate a comprehensive understanding the immunity landscape in tumor microenvironments [43] Moreover, future efforts in the development of single-cell sequencing technology will also emerge, which could focus on single-cell multimodal omics, such as genomic DNA–messenger RNA sequencing [44,45], genome and transcriptome sequencing [46], and single-cell methylome and transcriptome sequencing [47]. Currently, the areas of particular relevance to immuno-oncology based upon single-cell sequencing technology focus on the ability to track individual specific cell clones through paired sequencing of their cell receptor genes and high-dimensional single-cell spatial analysis [48]. High-dimensional single-cell sequencing technologies are likely to generate clinically relevant biomarker signatures in immuno-oncology and may be able to guide clinical decision making.

Conclusions

The number of publications related to single-cell sequencing technology has increased dramatically since 2014. The United States led absolute productivity ranks by contributing approximately 60% of the total publications, followed by China, the United Kingdom, and Germany. Similarly, collaboration networks consisted mainly of institutes and authors from the above-mentioned countries. Moreover, in terms of the top 10 productive journals, the most cited journals, and journals with the highest average citation, there were 20 journals calculated in total. And most of them were considered as top-journals (impact factor＞10, Q1).Single-cell sequencing technology has made a large impact in various fields of biology, with a noticeable increase in publications annually in the last 5 years. We believe that the field of immunology might be a future research hotspot.

Acknowledgments

We thank MogoEdit, a professional copyediting agency, for helping to improve the quality of language in our manuscript. This work was supported by National Key Research and Development Program of China (2017YFC1308805), Industry-University-Research Innovation Fund in Ministry of Education of the People’s Republic of China (2018A01013) and National Natural Science Foundation of China (grant 81871962).

Authors' Contributions

QW, KLY, ZZ, SW, and KWJ designed the study. KLY and ZW searched the database using the search strategy from JHT and LL. QW, ZW, KWJ, KLY, ZZ, and YJY screened the titles and abstracts to select articles. KLY, ZW, and YJY conducted analysis and mapping. All authors participated in the interpretation of the results. QW, KLY, SW, and KWJ wrote the manuscript with input from all authors. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Distribution of top 10 productive authors.
[DOCX File , 15 KB - jmir_v23i8e25789_app1.docx]

Multimedia Appendix 2
Distribution by top journals and its citations.
[DOCX File , 17 KB - jmir_v23i8e25789_app2.docx]

Multimedia Appendix 3
Top 36 references with strongest citation bursts.
[PNG File , 438 KB - jmir_v23i8e25789_app3.png]

Multimedia Appendix 4
Top 15 keywords with the strongest citation bursts.
[PNG File , 72 KB - jmir_v23i8e25789_app4.png]

Multimedia Appendix 5
Clusters of immunotherapy in 2019.
[PNG File , 677 KB - jmir_v23i8e25789_app5.png]

References

1. Bianconi E, Piovesan A, Facchin F, Beraudi A, Casadei R, Frabetti F, et al. An estimation of the number of cells in the human body. Ann Hum Biol 2013;40(6):463-471. [doi: 10.3109/03014460.2013.807878] [Medline: 23829164]
2. Wang Y, Navin NE. Advances and applications of single-cell sequencing technologies. Mol Cell 2015;58(4):598-609 [FREE Full text] [doi: 10.1016/j.molcel.2015.05.005] [Medline: 26000845]
3. Tang X, Huang Y, Lei J, Luo H, Zhu X. The single-cell sequencing: new developments and medical applications. Cell Biosci 2019;9:53 [FREE Full text] [doi: 10.1186/s13578-019-0314-y] [Medline: 31391919]
4. Wen L, Tang F. Boosting the power of single-cell analysis. Nat Biotechnol 2018;36(5):408-409. [doi: 10.1038/nbt.4131] [Medline: 29734314]
5. Stegle O, Teichmann SA, Marioni JC. Computational and analytical challenges in single-cell transcriptomics. Nat Rev Genet 2015;16(3):133-145. [doi: 10.1038/nrg3833] [Medline: 25628217]
6. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, et al. mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods 2009;6(5):377-382. [doi: 10.1038/nmeth.1315] [Medline: 19349980]
7. Dal Molin A, Di Camillo B. How to design a single-cell RNA-sequencing experiment: pitfalls, challenges and perspectives. Brief Bioinform 2019;20(4):1384-1394. [doi: 10.1093/bib/bby007] [Medline: 29394315]
8. Li J, Lin H, Hou R, Shen J, Li X, Xing J, et al. Multi-omics study in monozygotic twins confirm the contribution of de novo mutation to psoriasis. J Autoimmun 2020;106:102349 [FREE Full text] [doi: 10.1016/j.jaut.2019.102349] [Medline: 31629629]
9. Zhang Q, He Y, Luo N, Patel SJ, Han Y, Gao R, et al. Landscape and dynamics of single immune cells in hepatocellular carcinoma. Cell 2019 Oct 31;179(4):829-845.e20 [FREE Full text] [doi: 10.1016/j.cell.2019.10.003] [Medline: 31675496]
10. Wang Y, Niu Q, Zhang X, Liu L, Wang Y, Chen Y, et al. Exploring the effects of operational mode and microbial interactions on bacterial community assembly in a one-stage partial-nitritation anammox reactor using integrated multiomics. Microbiome 2019;7(1):122 [FREE Full text] [doi: 10.1186/s40168-019-0730-6] [Medline: 31462278]
11. Lake BB, Chen S, Sos BC, Fan J, Kaeser GE, Yung YC, et al. Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. Nat Biotechnol 2018;36(1):70-80 [FREE Full text] [doi: 10.1038/nbt.4038] [Medline: 29227469]
12. Rosenberg AB, Roco CM, Muscat RA, Kuchina A, Sample P, Yao Z, et al. Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. Science 2018;360(6385):176-182 [FREE Full text] [doi: 10.1126/science.aam8999] [Medline: 29545511]
13. Guo F, Li L, Li J, Wu X, Hu B, Zhu P, et al. Single-cell multi-omics sequencing of mouse early embryos and embryonic stem cells. Cell Res 2017;27(8):967-988 [FREE Full text] [doi: 10.1038/cr.2017.82] [Medline: 28621329]
14. Casasent AK, Schalck A, Gao R, Sei E, Long A, Pangburn W, et al. Multiclonal invasion in breast tumors identified by topographic single cell sequencing. Cell 2018;172(1-2):205-217 [FREE Full text] [doi: 10.1016/j.cell.2017.12.007] [Medline: 29307488]
15. Suvà ML, Tirosh I. Single-cell RNA sequencing in cancer: lessons learned and emerging challenges. Mol Cell 2019;75(1):7-12 [FREE Full text] [doi: 10.1016/j.molcel.2019.05.003] [Medline: 31299208]
16. Zhang X, Liu L. Applications of single cell RNA sequencing to research of stem cells. World J Stem Cells 2019;11(10):722-728 [FREE Full text] [doi: 10.4252/wjsc.v11.i10.722] [Medline: 31692946]
17. Ren X, Kang B, Zhang Z. Understanding tumor ecosystems by single-cell sequencing: promises and limitations. Genome Biol 2018;19(1):211 [FREE Full text] [doi: 10.1186/s13059-018-1593-z] [Medline: 30509292]
18. Kiselev VY, Andrews TS, Hemberg M. Challenges in unsupervised clustering of single-cell RNA-seq data. Nat Rev Genet 2019 May;20(5):273-282. [doi: 10.1038/s41576-018-0088-9] [Medline: 30617341]
19. Agarwal A, Durairajanayagam D, Tatagari S, Esteves SC, Harlev A, Henkel R, et al. Bibliometrics: tracking research impact by selecting the appropriate metrics. Asian J Androl 2016;18(2):296-309 [FREE Full text] [doi: 10.4103/1008-682X.171582] [Medline: 26806079]
20. Keathley-Herring H, Van Aken E, Gonzalez-Aleu F, Deschamps F, Letens G, Orlandini PC. Assessing the maturity of a research area: bibliometric review and proposed framework. Scientometrics 2016 Aug 13;109(2):927-951. [doi: 10.1007/s11192-016-2096-x]
21. Chen C. CiteSpace II: detecting and visualizing emerging trends and transient patterns in scientific literature. J Am Soc Inf Sci 2006 Feb 01;57(3):359-377. [doi: 10.1002/asi.20317]

22. van Eck NJ, Waltman L. Citation-based clustering of publications using CitNetExplorer and VOSviewer. Scientometrics 2017;111(2):1053-1070 [FREE Full text] [doi: 10.1007/s11192-017-2300-7] [Medline: 28490825]
23. Devos P, Ménard J. Trends in worldwide research in hypertension over the period 1999-2018: a bibliometric study. Hypertension 2020;76(5):1649-1655 [FREE Full text] [doi: 10.1161/HYPERTENSIONAHA.120.15711] [Medline: 32862706]
24. van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics 2010 Aug;84(2):523-538 [FREE Full text] [doi: 10.1007/s11192-009-0146-3] [Medline: 20585380]
25. van Eck NJ, Waltman L. Text mining and visualization using VOSviewer. Arxiv 2011 Sep 09 [FREE Full text]
26. Ruiz Sebastián C, O'Ryan C. Single-cell sequencing of dinoflagellate (Dinophyceae) nuclear ribosomal genes. Mol Ecol Notes 2001;1(4):329-331. [doi: 10.1046/j.1471-8278.2001.00084.x]
27. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 2012;19(5):455-477 [FREE Full text] [doi: 10.1089/cmb.2012.0021] [Medline: 22506599]
28. Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, et al. Tumour evolution inferred by single-cell sequencing. Nature 2011;472(7341):90-94 [FREE Full text] [doi: 10.1038/nature09807] [Medline: 21399628]
29. Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H, et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science 2014;344(6190):1396-1401 [FREE Full text] [doi: 10.1126/science.1254257] [Medline: 24925914]
30. Tirosh I, Izar B, Prakadan SM, Wadsworth MH, Treacy D, Trombetta JJ, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science 2016;352(6282):189-196 [FREE Full text] [doi: 10.1126/science.aad0501] [Medline: 27124452]
31. Zeisel A, Muñoz-Manchado AB, Codeluppi S, Lönnerberg P, La Manno G, Juréus A, et al. Brain structure. cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science 2015;347(6226):1138-1142 [FREE Full text] [doi: 10.1126/science.aaa1934] [Medline: 25700174]
32. Kolodziejczyk AA, Kim JK, Svensson V, Marioni JC, Teichmann SA. The technology and biology of single-cell RNA sequencing. Mol Cell 2015;58(4):610-620 [FREE Full text] [doi: 10.1016/j.molcel.2015.04.005] [Medline: 26000846]
33. Li A, Tang X. Foundation support for studies on single-cell sequencingits application in USA, Britain and China. Chin J Med Lib Info Sci 2017;26(7):44-51. [doi: 10.3969/j.issn.1671-3982.2017.09.008]
34. Metzker ML. Emerging technologies in DNA sequencing. Genome Res 2005;15(12):1767-1776 [FREE Full text] [doi: 10.1101/gr.3770505] [Medline: 16339375]
35. Xu X, Hou Y, Yin X, Bao L, Tang A, Song L, et al. Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. Cell 2012;148(5):886-895 [FREE Full text] [doi: 10.1016/j.cell.2012.02.025] [Medline: 22385958]
36. Wang J, Fan HC, Behr B, Quake SR. Genome-wide single-cell analysis of recombination activity and de novo mutation rates in human sperm. Cell 2012;150(2):402-412 [FREE Full text] [doi: 10.1016/j.cell.2012.06.030] [Medline: 22817899]
37. Evrony GD, Cai X, Lee E, Hills LB, Elhosary PC, Lehmann HS, et al. Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain. Cell 2012;151(3):483-496 [FREE Full text] [doi: 10.1016/j.cell.2012.09.035] [Medline: 23101622]
38. Ramsköld D, Luo S, Wang Y, Li R, Deng Q, Faridani OR, et al. Full-length mRNA-seq from single-cell levels of RNA and individual circulating tumor cells. Nat Biotechnol 2012;30(8):777-782 [FREE Full text] [doi: 10.1038/nbt.2282] [Medline: 22820318]
39. Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J, et al. Insights into the phylogeny and coding potential of microbial dark matter. Nature 2013;499(7459):431-437. [doi: 10.1038/nature12352] [Medline: 23851394]
40. Shalek AK, Satija R, Adiconis X, Gertner RS, Gaublomme JT, Raychowdhury R, et al. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. Nature 2013;498(7453):236-240 [FREE Full text] [doi: 10.1038/nature12172] [Medline: 23685454]
41. Nagano T, Lubling Y, Stevens TJ, Schoenfelder S, Yaffe E, Dean W, et al. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. Nature 2013;502(7469):59-64 [FREE Full text] [doi: 10.1038/nature12593] [Medline: 24067610]
42. Peng C, He M, Cutrona SL, Kiefe CI, Liu F, Wang Z. Theme trends and knowledge structure on mobile health apps: bibliometric analysis. JMIR Mhealth Uhealth 2020;8(7):e18212 [FREE Full text] [doi: 10.2196/18212] [Medline: 32716312]
43. Zheng Y, Chen Z, Han Y, Han L, Zou X, Zhou B, et al. Immune suppressive landscape in the human esophageal squamous cell carcinoma microenvironment. Nat Commun 2020;11(1):6268 [FREE Full text] [doi: 10.1038/s41467-020-20019-0] [Medline: 33293583]
44. Zhu C, Preissl S, Ren B. Single-cell multimodal omics: the power of many. Nat Methods 2020;17(1):11-14. [doi: 10.1038/s41592-019-0691-5] [Medline: 31907462]
45. Dey SS, Kester L, Spanjaard B, Bienko M, van Oudenaarden A. Integrated genome and transcriptome sequencing of the same cell. Nat Biotechnol 2015;33(3):285-289 [FREE Full text] [doi: 10.1038/nbt.3129] [Medline: 25599178]
46. Macaulay IC, Haerty W, Kumar P, Li YI, Hu TX, Teng MJ, et al. Nat Methods 2015;12(6):519-522. [doi: 10.1038/nmeth.3370] [Medline: 25915121]

47. Angermueller C, Clark SJ, Lee HJ, Macaulay IC, Teng MJ, Hu TX, et al. Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. Nat Methods 2016;13(3):229-232 [FREE Full text] [doi: 10.1038/nmeth.3728] [Medline: 26752769]
48. Gohil SH, Iorgulescu JB, Braun DA, Keskin DB, Livak KJ. Applying high-dimensional single-cell technologies to the analysis of cancer immunotherapy. Nat Rev Clin Oncol 2021 Apr;18(4):244-256. [doi: 10.1038/s41571-020-00449-x] [Medline: 33277626]

Abbreviations

WoSCC: Web of Science Core Collection

Edited by R Kukafka; submitted 16.11.20; peer-reviewed by X Wang, L Ge; comments to author 09.12.20; revised version received 22.12.20; accepted 13.05.21; published 10.08.21.

Please cite as:
Wang Q, Yang KL, Zhang Z, Wang Z, Li C, Li L, Tian JH, Ye YJ, Wang S, Jiang KW
Characterization of Global Research Trends and Prospects on Single-Cell Sequencing Technology: Bibliometric Analysis
J Med Internet Res 2021;23(8):e25789
URL: https://www.jmir.org/2021/8/e25789
doi:10.2196/25789
PMID:34014832

©Quan Wang, Ke-Lu Yang, Zhen Zhang, Zhu Wang, Chen Li, Lun Li, Jin-Hui Tian, Ying-Jiang Ye, Shan Wang, Ke-Wei Jiang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Corrigenda and Addenda

Correction: The State of Evidence in Patient Portals: Umbrella Review

Marcy G Antonio[1], BSc, MPH; Olga Petrovskaya[2], RN, PhD; Francis Lau[1], PhD, FCAHS, CTSS

[1]University of Victoria, Victoria, BC, Canada

[2]University of Alberta, Edmonton, AB, Canada

Corresponding Author:
Marcy G Antonio, BSc, MPH
University of Victoria
PO Box 1700 STN CSC
Victoria, BC, V8W 2Y2
Canada
Phone: 1 7783501089
Email: mantonio@uvic.ca

Related Article:

Correction of: https://www.jmir.org/2020/11/e23851

(J Med Internet Res 2021;23(8):e32421) doi:10.2196/32421

In “The State of Evidence in Patient Portals: Umbrella Review” (J Med Internet Res 2020;22(11):e23851) the authors added a clarification.

In the section “Rating the Umbrella Review Evidence,” two sentences have been added to the text to clarify that this work was not done under the guidance of the GRADE working group but was conceptualized independently of the group. In the originally published paper, the first paragraph of this section appeared as follows:

> *As the concluding step, 2 researchers independently assessed the strength of evidence for quantitative umbrella review finding statements and the confidence in the evidence for qualitative umbrella review finding statements. For this purpose, we developed meta-level umbrella review tools, GRADE-UR (Grading of Strength of Evidence for Quantitative Research at the Level of an Umbrella Review) and CERQual-UR (Grading of Confidence in the Evidence of Qualitative Research at the Level of an Umbrella Review), by applying a voting-counting method [38] and adapting GRADE [39-41] and CERQual [42-44] SR evaluation tools.*

This has been corrected to:

> *As the concluding step, 2 researchers independently assessed the strength of evidence for quantitative umbrella review finding statements and the confidence in the evidence for qualitative umbrella review finding statements. For this purpose, we developed meta-level umbrella review tools, GRADE-UR (Grading of Strength of Evidence for Quantitative Research at the Level of an Umbrella Review) and CERQual-UR (Grading of Confidence in the Evidence of Qualitative Research at the Level of an Umbrella Review), by applying a voting-counting method [38] and adapting GRADE [39-41] and CERQual [42-44] SR evaluation tools. The methodological approach was conceptualized independent from the GRADE working group. The GRADE-UR and CERQual-UR acronym was created by the authors of this paper to reflect an adaptation of GRADE and CERQual.*

The correction will appear in the online version of the paper on the JMIR Publications website on August 16, 2021, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 27.07.21; this is a non–peer-reviewed article;accepted 02.08.21; published 16.08.21.

Please cite as:
Antonio MG, Petrovskaya O, Lau F
Correction: The State of Evidence in Patient Portals: Umbrella Review
J Med Internet Res 2021;23(8):e32421
URL: https://www.jmir.org/2021/8/e32421
doi:10.2196/32421
PMID:34398799

©Marcy G Antonio, Olga Petrovskaya, Francis Lau. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 16.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Evaluating Epidemiological Risk by Using Open Contact Tracing Data: Correlational Study

Stefano Piotto[1,2], PhD; Luigi Di Biasi[3], MSc; Francesco Marrafino[1], PhD; Simona Concilio[1,2], PhD

[1]Department of Pharmacy, University of Salerno, Fisciano, Italy
[2]Bionam Research Center for Biomaterials, University of Salerno, Fisciano, Italy
[3]Department of Computer Sciences, University of Salerno, Fisciano, Italy

Corresponding Author:
Stefano Piotto, PhD
Department of Pharmacy
University of Salerno
Via Giovanni Paolo II, 132
Fisciano, 84084
Italy
Phone: 39 3204230068
Email: piotto@unisa.it

Abstract

Background: During the 2020s, there has been extensive debate about the possibility of using contact tracing (CT) to contain the SARS-CoV-2 pandemic, and concerns have been raised about data security and privacy. Little has been said about the effectiveness of CT. In this paper, we present a real data analysis of a CT experiment that was conducted in Italy for 8 months and involved more than 100,000 CT app users.

Objective: We aimed to discuss the technical and health aspects of using a centralized approach. We also aimed to show the correlation between the acquired contact data and the number of SARS-CoV-2–positive cases. Finally, we aimed to analyze CT data to define population behaviors and show the potential applications of real CT data.

Methods: We collected, analyzed, and evaluated CT data on the duration, persistence, and frequency of contacts over several months of observation. A statistical test was conducted to determine whether there was a correlation between indices of behavior that were calculated from the data and the number of new SARS-CoV-2 infections in the population (new SARS-CoV-2–positive cases).

Results: We found evidence of a correlation between a weighted measure of contacts and the number of new SARS-CoV-2–positive cases (Pearson coefficient=0.86), thereby paving the road to better and more accurate data analyses and spread predictions.

Conclusions: Our data have been used to determine the most relevant epidemiological parameters and can be used to develop an agent-based system for simulating the effects of restrictions and vaccinations. Further, we demonstrated our system's ability to identify the physical locations where the probability of infection is the highest. All the data we collected are available to the scientific community for further analysis.

(J Med Internet Res 2021;23(8):e28947) doi:10.2196/28947

KEYWORDS

SARS-CoV-2; COVID-19; contact tracing; Bluetooth Low Energy; transmission dynamics; infection spread; mobile apps; mHealth; digital apps; mobile phone

Introduction

In China, during December 2019, SARS-CoV-2 was identified as a novel beta coronavirus. At the time of writing this paper (December 2020), SARS-CoV-2 has caused almost 60 million confirmed human infections worldwide and more than 1 million deaths since its discovery [1,2]. The disease caused by SARS-CoV-2 is called COVID-19, and the disease was declared a global pandemic on March 11, 2020 [3]. Containment measures are the first and most crucial step for rapidly halting an outbreak that could otherwise become an epidemic or even turn into a pandemic, such as the COVID-19 outbreak [4].

Notable examples of disease epidemics with a high occurrence of superspreading events (SSEs) are the SARS-CoV (severe acute respiratory syndrome coronavirus; 2002-2003) and MERS-CoV (Middle East respiratory syndrome coronavirus; since 2013) epidemics [5-9]. The basic reproduction number (R_0) is a key measure of transmissibility. It is defined as the number of infected contacts that 1 infected individual generates on average during their infectious period. An R_0 value of >1 means that a virus will continue its propagation among susceptible hosts. In contrast, an R_0 of <1 means that it is certain that epidemic spread will stop [10,11]. The SARS-CoV and MERS-CoV have an R_0 of around 3 [12]. For SARS-CoV-2, the estimated R_0 ranges between 2 and 3 [9,13]. However, it is unknown as to what extent SSEs are involved in the spread of SARS-CoV-2 infection.

Lockdown was the most widespread pandemic containment response, and it was introduced at different levels by most affected countries. As already predicted by mathematical models [14] and proven by trends that were updated at the time of writing this paper, the contagion's spread resumed rapidly when lockdown countermeasures were lifted. Rapid and automatic contact tracing (CT) is an essential intervention for contagion containment [15-19]; however, user localization poses a privacy risk and reduces compliance rates [20]. According to the World Health Organization, CT involves the following three steps: the identification of a contact (identifying those that a confirmed positive patient had contact with based on the transmission modalities of the pathogen of interest), the listing of contacts (keeping a record of individuals who possibly had contact with infected patients and informing these individuals), and contact follow-up [21]. CT has a dual purpose—treating people who have possibly been exposed to infectious diseases and stopping the transmission chain to contain an epidemic. Due to the prevalence of smartphones, CT has the potential to become a powerful intervention; the vast majority of smartphone users carry their smartphone devices with them throughout the day, and smartphones can generate detailed GPS location information. However, due to the availability of users' location data, there is growing concern about the infringement of an individual's right to privacy. An alternative is using other contact monitoring technologies that are based on proximity assessments rather than those based on location information [22]. It is important to note that this study does not constitute an endorsement or rejection of CT based on potential data security risks or privacy limitations. This study intends to assess whether and to what extent the acquisition of contact data helps with assessing the spread of SARS-CoV-2.

Technologies such as Bluetooth Low Energy allow for the evaluation of the distance between users without locating them and thus help with addressing the privacy issue. The number of CT apps that have been introduced since the beginning of the SARS-CoV-2 pandemic is considerable [23,24] and reflects governments' interest in automating the tracing of people who have had recent contact with individuals who tested positive for COVID-19. An app that uses a centralized approach was developed by the academic spin-off company of the University of Salerno—SoftMining (SM). The app [25] was supported by government agencies such as the Campania Region and was validated by more than 120,000 users; the app had peaks of more than 15,000 active daily users.

CT is a fundamental intervention for acquiring population data, which show how different population groups can behave differently. Such behaviors result in different risks of infection among group members. In Multimedia Appendix 1, we describe how CT data were acquired via the Bluetooth Low Energy technology of the SM-COVID-19 app and how data were clustered to obtain different mobility and behavior groups. In this paper, we discuss how we used Italian National Institute of Health data on contagion trends in Italy [26] to estimate a more precise number of SARS-CoV-2–positive cases that was less influenced by the number of tests performed on the population. In addition, we show the link between the acquired CT data and the number of new SARS-CoV-2–positive cases. This allowed us to define an epidemiological risk function that was based on the number of, frequency of, and distance between contacts. The risk function expresses the probability that an individual will become ill as a function of their age within a given period of time. This study aims to evaluate whether the use of CT can support the containment of an epidemic. The data acquired from CT were analyzed and correlated with data on the progression of SARS-CoV-2 infection.

This study was not conducted for commercial purposes; it was conducted for the purposes of academic research and aims to make CT data available to the scientific community for future research.

Methods

CT Data Acquisition

During the CT phase, the SM-COVID-19 app analyzed the environment and, at regular intervals, sent data on the duration of a contact and the instantaneous and average distances (over the time) of a contact to the server. App users could voluntarily decide to share location data as well. If they did, the server also received latitude, longitude, precision, and smartphone provider data. We provide the full description of the data acquisition procedures in Multimedia Appendices 1 and 2. The developed technologies allowed for high precision in distance calculations (less than 0.5 m under optimal conditions and after device calibration) and were implemented via the SM-COVID-19 app, which is available on Android and iOS smartphones (via TestFlight; Apple Inc). Daily data were anonymized and saved for further use, as described in Multimedia Appendix 3, in accordance with the General Data Protection Regulation. Anonymity was also guaranteed when the GPS localization function was enabled, as data were stored randomly in the database; the database did not present an individual user's location in a precise way. The app only used random 128-bit proximity IDs, and only the user's device kept track of the device IDs. The app's functions were conducted and maintained with a back-end server, on which arbitrary identifiers were stored. Users could not be identified directly with app data, as only the app's random identifiers were stored on the server.

Social Mobility Analysis

The data set obtained from the SM-COVID-19 app in the period of April to November 2020 was analyzed. The data set's structure is described in Multimedia Appendix 2. Reported data from August 1 to August 30, 2020, were obtained to analyze mobility data from a period when no lockdown measures were in place. Such data are useful for tracking movements in real situations. We removed users with less than 15 days of activity from our analysis to exclude users who may have deactivated the app. The cleaned data set was clustered. Before the clustering process, the t-distributed stochastic neighbor embedding machine learning algorithm was applied to the data set to reduce its dimensionality to 2. The clustering was carried out by using the Ward linkage method. This method allows the user to select the number of clusters arbitrarily. We analyzed the distribution of data for different numbers of clusters (2-10 clusters); the optimal distribution was obtained with 5 clusters. The average number of daily contacts and the SDs for the clusters are reported in Table 1. SDs were high, since every cluster had many users with 0-contact days among those with low- and high-contact days. As shown in Table 1, the population was divided into clusters of approximately the same size. However, cluster 5 was larger and included users who had a larger number of contacts. This cluster accounted for the population with the highest number of contacts and included users with the highest number of contacts and the highest mobility.

Table 1. The cluster data of active users for the period of August 1 to August 30, 2020.

Cluster number	Number of daily contacts based on Bluetooth Low Energy technology, mean (SD)[a]	Percentage of active users
1	23.40 (38.55)	14
2	12.05 (22.62)	19
3	41.95 (75.79)	17
4	69.91 (103.76)	20
5	121.48 (145.05)	30

[a] The average number of daily contacts for each cluster and SDs were calculated based on all cluster data (ie, from days 1 to 30).

Data Availability

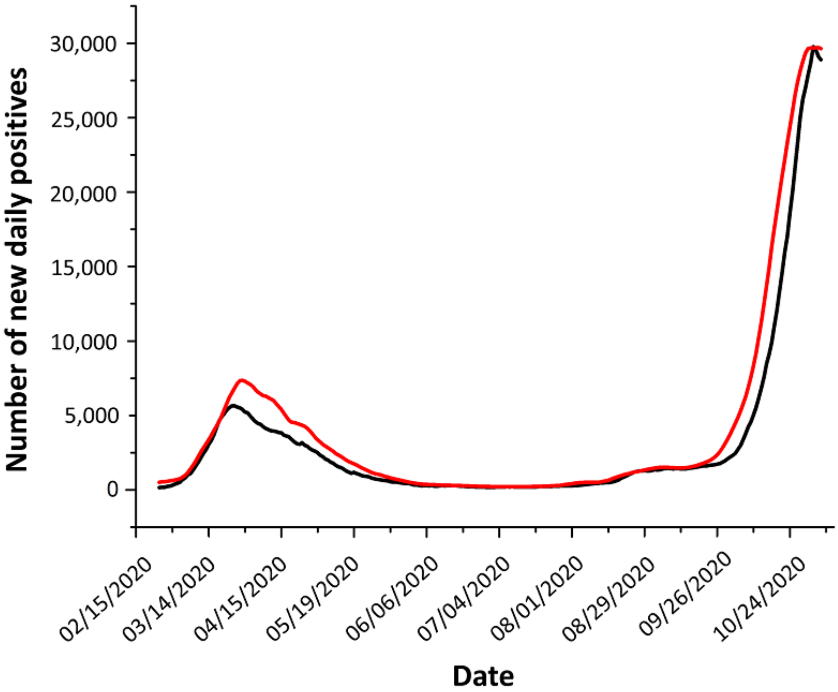
All data can be made available upon request from the authors or the SM-COVID-19 team [27].

Results

Statistical Analysis and Estimates of the Real Number of SARS-CoV-2–Positive Cases

For our statistical analysis, we relied on official data on the daily SARS-CoV-2–related trends in Italy, which were released by the Italian National Institute of Health and aggregated by the Department of Civil Protection of the Presidency of the Council of Ministers [17]. We estimated the possible number of real infections that may have occurred during the epidemic in Italy. We obtained the daily number of newly performed tests based on the total number of tests performed. This was calculated by using equation 1 in Multimedia Appendix 4. The method for estimating the number of new daily SARS-CoV-2–positive cases is detailed in Multimedia Appendix 4. We performed data smoothing via sliding-window averaging to reduce each day's variability, which was the result of the cumulative regional data's intrinsic variability. The SARS-CoV-2–related trends over a given period were roughly linear; there were no sudden peaks. Additionally, the averaging process performed allowed us to smoothen the curves, which were in line with these trends. Equation 2 in Multimedia Appendix 4 was used to define the ratio between the number of daily tests and the number of daily reported SARS-CoV-2–positive cases. The estimated number of new SARS-CoV-2–positive cases ($EP_{[k]}$) for each day was calculated with equation 3 in Multimedia Appendix 4. With our method, we estimated a correction for the number of real SARS-CoV-2–positive cases that occurred during the pandemic period. We observed that around 224,000 cases were not diagnosed, and of these cases, nearly around 81,000 were missed in the period of March to May 2020. The difference between the official number of cases and the estimated number of new SARS-CoV-2–positive cases is shown in Figure 1.

Figure 1. Comparison of the official number of daily new SARS-CoV-2–positive cases reported by the ISS (black line) and the estimated number of daily new SARS-CoV-2–positive cases (red line). The difference was higher during the initial phases of the pandemic. ISS: Istituto Superiore di Sanità (Italian Superior Institute of Health).



Correlation Between CT Data and Contagion Trends

The correct number of daily SARS-CoV-2–positive cases was calculated to perform correlation analyses with the data obtained from CT. The data distributed by the ISS, due to how the data were structured, showed considerable fluctuations based on the number of tests performed. It was also possible to observe a weekly trend in the number of SARS-CoV-2–positive cases recorded due to the reduced number of tests performed during weekends. Such data therefore presented fluctuations that could alter the analysis. Data smoothing via sliding-window averaging also provided an additional element for alleviating the issue with fluctuations.

We then examined whether the contact index (CI) and the alpha index (α) correlated with the number of daily new SARS-CoV-2–positive cases. These two parameters are indices of effective contacts and account for the distance between two users who come into contact with each other and the contact's duration. These parameters and the related equations are described in detail in Multimedia Appendix 5 [4,28,29]. These parameters were necessary, since not all of the contacts recorded by the app involved people who could effectively transmit the virus. CI_k is a value that indicates a user's risk of infection on day k based on the number of effective contacts that the user had on the same day. CI_k was calculated with equation 4 in Multimedia Appendix 5 [4,28,29]. α_k is a risk index, and it is based on data from the previous k−14 days (excluding day k). α_k reflects a user's behavior. The optimization of these parameters will be the subject of future studies.

The SM-COVID-19 data set lists the CI and α values for each day and every user. Therefore, to evaluate daily trends, we calculated the total CI and α values for each day (k) by summing each individual users' values. As such, it was possible to evaluate the trends for CI and α values and exclude users who deactivated the app for a given period. The values were smoothed by using a sliding window of 7 days. In Figure 2, we show the temporal evolution of CI values over 160 days. For visualization, in Figure 2, we report the logarithm of the number of new SARS-CoV-2–positive cases. There is an evident, rough correlation between the CI and the number of new SARS-CoV-2–positive cases. For each CI_k and α_k value, we calculated the Pearson correlation coefficient based on the estimated number of SARS-CoV-2–positive cases to assess how the number of contacts varied before and after a confirmation of COVID-19 positivity. It was very interesting to note that the correlation coefficient for CI_k reached its maximum at k+7 days. The high correlations observed in the subsequent days correlated with SARS-CoV-2 incubation times, and COVID-19 positivity occurred in the days following an effective contact. The α_k value reached its maximum at k+5 days. The differences between the α and CI values' correlation coefficients (ie, their correlation with the number of new SARS-CoV-2–positive cases) were attributable to the different calculation methods that were used for the two parameters, as the α value accounts for the risk of infection in the 14 days before day k. The correlation between CI values and the number of new SARS-CoV-2–positive cases is shown in Figure 3. We reported the correlation data that corresponded to the period of June to October 2020 because of the high availability of more consistent CT data. This correlation was also monitored for the previous studied period (March to May 2020) to confirm that the obtained values were not the result of artifacts or autocorrelations.

Figure 2. Temporal evolution of the CI values (black line) and the logarithm of the number of new SARS-CoV-2–positive cases (red line) during a 160-day period. CI: contact index.

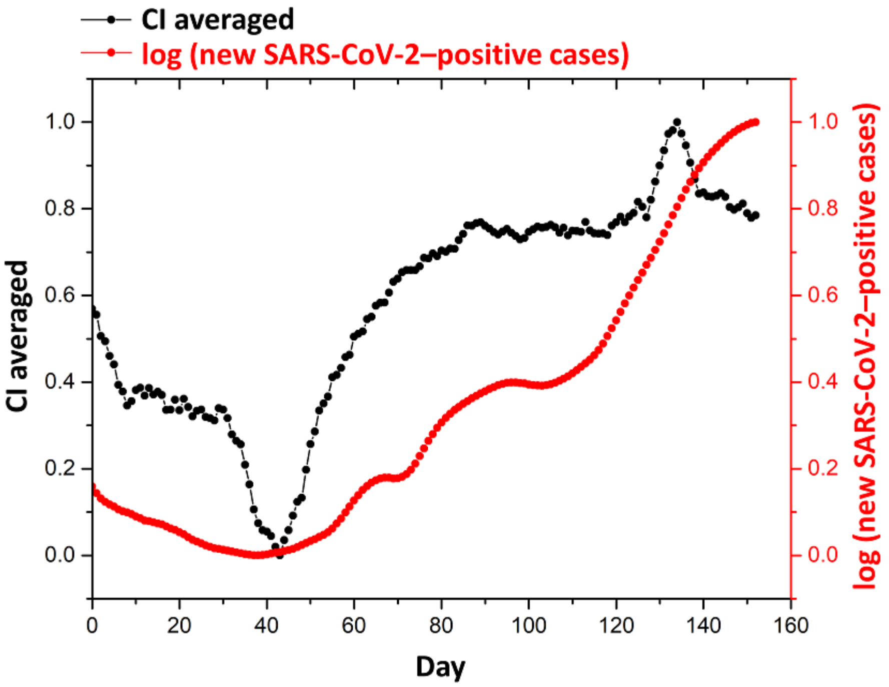
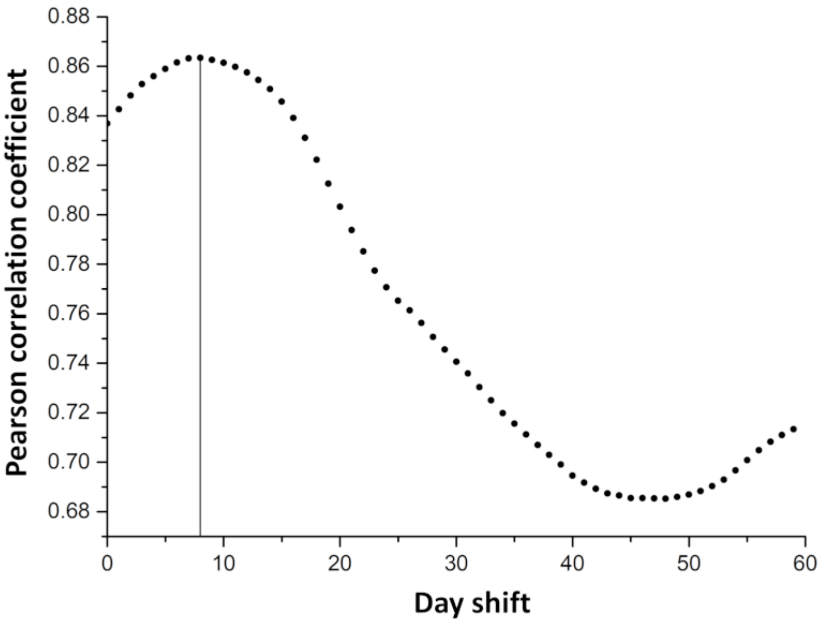


Figure 3. Pearson correlation analysis between CIk and the number of new SARS-CoV-2–positive cases with a time shift of 0 days to 60 days for the period of June 1 to October 31, 2020. The highest correlation value was observed at k+7 days. CI: contact index.



Discussion

The analysis of the collected data allowed us to determine the aspects of CT that are essential for the evaluation of the progression of the SARS-CoV-2 pandemic. These essential aspects were identified via the estimation of the real number of new SARS-CoV-2–positive cases and the correlation of the number and frequency of contacts with the probability of infection.

Estimation of the Total Number of People Who Tested Positive for SARS-CoV-2

At the beginning of the pandemic in Italy, during the period of March to May 2020, the substantial underestimation of the total number of people who tested positive for SARS-CoV-2 in Italy was a likely scenario. This was undoubtedly due to the reduced number of tests that were performed during the first phase of the SARS-CoV-2 pandemic and the lack of an adequate response for tracing infections. One method for estimating a realistic number of SARS-CoV-2–positive cases is to use the ratio between the number of tests carried out and the number of SARS-CoV-2–positive cases detected every day. We chose this ratio because as the number of tests carried out increases, this number eventually plateaus. These data are collected throughout the country and are therefore subject to regional and local variability. It has been assumed that the ratio between the number of positive cases and the number of tests performed varies slowly over time in the absence of hospitalization problems. This ratio has been used to estimate the actual number of SARS-CoV-2–positive cases, which is always greater than or equal to the official number of cases. As shown in Figure 1, the difference between the official number of daily new SARS-CoV-2–positive cases and the estimated number of cases was higher during the initial phases of the pandemic (ie, during the period of March to May 2020). During this period, according to our analysis, at least 81,000 patients with SARS-CoV-2 infection were not diagnosed with COVID-19. As already mentioned, calculating the real number of new SARS-CoV-2–positive cases was necessary because the data provided by the Istituto Superiore di Sanità (Italian Superior Institute of Health) varied according to the number of tests performed each day. In the initial stages of the pandemic, the number of tests was remarkably low due to the lack of adequate diagnostic tools.

Ethical and Practical Issues of CT Apps

CT apps have generated much discussion, particularly discussions regarding privacy and such apps’ susceptibility to attacks. Considerations of data security and possible privacy violations are certainly essential elements and have resulted in the creation of numerous solutions that have been adopted at the national level. This paper does not aim not to take a position on the security and privacy of CT apps, although the developers of SM-COVID-19 have considered these aspects. Rather, we are concerned with assessing whether CT apps, that is, those that can be developed based on currently available technology, can impact communities' health. Several apps have been adopted at a national level by multiple countries. However, during our research, we did not find any information on the availability of data collected by these apps. CT data provide useful information on various aspects of the SARS-CoV-2 pandemic (eg, the pandemic course) and the behavior and mobility of app users, thereby allowing researchers to map the frequency of contacts and identify high-risk areas. Our CT data set allowed us to analyze data and identify different classes of behavior among the population.

The SM-COVID-19 app uses a centralized model [23,24]. However, despite using a centralized model, users' privacy is completely protected via anonymization, as per the General Data Protection Regulation. The advantage of using a centralized model is that data stored on the server can be anonymized via aggregation and used by public authorities as a source of important aggregate information about the number of contacts in the population, the app's effectiveness in tracing and alerting contacts, and the aggregate number of people who could potentially develop symptoms. Unlike a decentralized model, a centralized model provides access to CT data, thereby making these data available for analysis and the improvement of epidemiological models. As already stated by Ferretti et al [19], the control of the SARS-CoV-2 epidemic via manual CT is impossible, as CT introduces a time lag resulting from the need to notify individuals about having contact with infected individuals. Such lag exacerbates the spread the infection, which is already remarkable given the infectivity of SARS-CoV-2 and the high percentage of transmission by presymptomatic individuals. The use of this app model, in which individuals are immediately notified about having contact with people who tested positive for SARS-CoV-2, would be sufficient for stopping the epidemic if the app is used by an adequate number of people [30] and would provide valuable data for creating accurate and valid predictive and epidemiological models. The choice of using a centralized model allows for the reconstruction of the chains of contagion transmission and the rapid propagation of risk indices (calculated with mathematical models)—operations that are difficult to implement when tracing data are only kept on devices.

By using data from August 2020, during which no lockdown measures or restrictions on mobility were in place and only partial restrictions were placed on gatherings, it was possible to identify 5 different behavior classes (or mobility classes). Table 1 shows the data from the clustering process. The five groups had approximately the same population size except for cluster 5, which had the largest number of people and included individuals with the highest mobility. The high amount of deviation in cluster 5 shows how users in this class alternated between experiencing days with 0 contacts (ie, no mobility; eg, days when they could be working from home) and experiencing days with a very high number of contacts (eg, due to a commute or due to work involving contact with the public). From these clusters, it is impossible to define the reasons behind a given number of contacts, but this is irrelevant as long as similar behaviors are present among the users belonging to a certain cluster. However, this clustering process provided interesting insights; it showed that there are classes of people with very low mobility (eg, older people) and classes of people with high mobility who experience a high number of contacts (eg, working in a hospital, supermarket, etc). This information can be even more useful when using a localized approach, such as using GPS data, as such data would help with providing more appropriate definitions for categories. The contacts registered by the app allowed us to trace the frequency of contacts and the trend in the number of contacts for a given period, a single user, a cluster, or the whole data set.

Correlation Between CT and the Total Number of New SARS-CoV-2–Positive Cases

CT data correlated with the growth in the number of new SARS-CoV-2–positive cases, and the highest correlation was observed 5 to 7 days after day k. This observation is in line with the hypothesis that an increase in the number of contacts is linked to an increased risk of infection. The most interesting element of the correlation is the time gap. The differences in the correlation values were probably related to the incubation period of SARS-CoV-2. Consequently, a contact that occurs on day k will not result in COVID-19 positivity on day k but on day k+n. This time gap is in line with the estimated incubation time for SARS-CoV-2 [4,28], and our analysis shows the effectiveness of using CT data to predict the number of new SARS-CoV-2–positive cases. This high correlation means that CT data can be used to develop new and more accurate epidemiological models and predictive tools.

Although a distributed approach that involves the use of a central advertising server makes it possible to alert individuals in direct contacts (the first contact between a newly infected individual and another person) about an eventual infection, flooding operations are necessary on CT networks to warn individuals about contacts of level 2 or higher. The decentralized model provides only 1 degree of separation from a CT app user who tested positive for COVID-19 (user A). To obtain data on a longer chain of contacts, which would have a decreasing risk gradient, it would be necessary for user B (a user in user A's contact chain) to publish their identifier so that user C (a user who had contact with user B but not with user A) is alerted. This could prove particularly dangerous when an asymptomatic or low-symptomatic individual who has not been tested for SARS-CoV-2 infection could infect another person and even cause another person's death. [31] In such a situation, decentralized CT would fail. On the other hand, the centralized model allows for the instant tracing of all contacts, regardless of the degree of separation. This would result in the more effective containment of the contagion, since all individuals in a contact chain that are deemed to be at risk for infection would be notified immediately about the danger. In this model, voluntary data input by individuals involved in first-degree contacts for informing those involved in second-degree contacts would not be required whenever the former was notified about having contact with a person who tested positive for COVID-19. Similar conclusions were reached by Aleta et al [30], who proved the effectiveness of using an automatic and extensive CT system to contain the spread of SARS-CoV-2 when lockdown measures are lifted. The work of Aleta et al [30] confirmed the usefulness of CT data collected from the population and provided an excellent basis for improving predictions and reducing the social and economic impact of SARS-CoV-2 prior to the effective vaccination of the entire population. At the time of writing this paper, we did not find any other available data sets with real CT data.

Geolocalization

CT data can be beneficial for evaluating SARS-CoV-2 propagation data. The data set that was made available by the app is particularly interesting because, due to its structure, it can be used as the basis for tracing SSEs. SSEs are generally defined as outbreaks in which a small number of individuals infect a large number of secondary individuals (ie, well-above the expected average number of individuals) [32]. The CT data that allowed us to define behavioral clusters for the population can also help with determining the SARS-CoV-2 pandemic's potential for generating SSEs. Although lower than those of the SARS-CoV and MERS-CoV pandemics, the SARS-CoV-2 pandemic's potential for generating SSEs is significant. In the absence of interventions such as social distancing, this potential would be even more significant. When developing disease control measures, people should focus on the rapid CT and quarantining of infected individuals and policies for physical distancing or targeted shutdowns to prevent the occurrence of SSEs. Having the ability to predict a pandemic's potential for generating SSEs would be vital in preventing outbreaks, and it would considerably reduce a contagion's overall R_0 value. The use of GPS data that are made anonymous with an appropriate protocol would enable researchers to use a rapid localized approach to significantly reducing the risk of contagion spread in certain areas and act in a targeted and localized manner. This type of information can prove very useful for planning the possible containment of a contagion in defined areas. The tests we performed that used GPS data showed the potential of this approach. For these tests, CT data that were acquired during the lockdown period (April 14 to May 3, 2020) from SM-COVID-19 users who had explicitly activated GPS tracing and whose GPS coordinates included the Campania Region were used (Figure 4). The simulations showed that a higher number of alerts were generated in locations that corresponded to the outbreaks that occurred during the lockdown (Figure 5). This type of voluntarily provided information can be a handy tool for confining and preventing contagion spread.

Figure 4. A map showing contact tracing app users' GPS locations on September 10, 2020. These data were used for the tests.

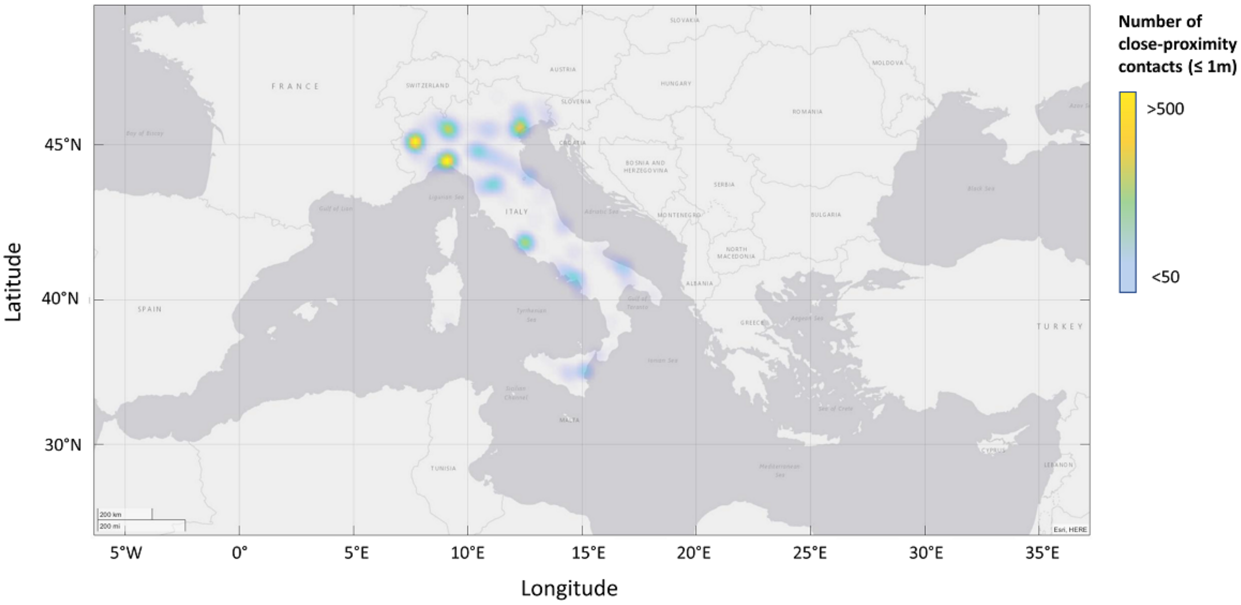
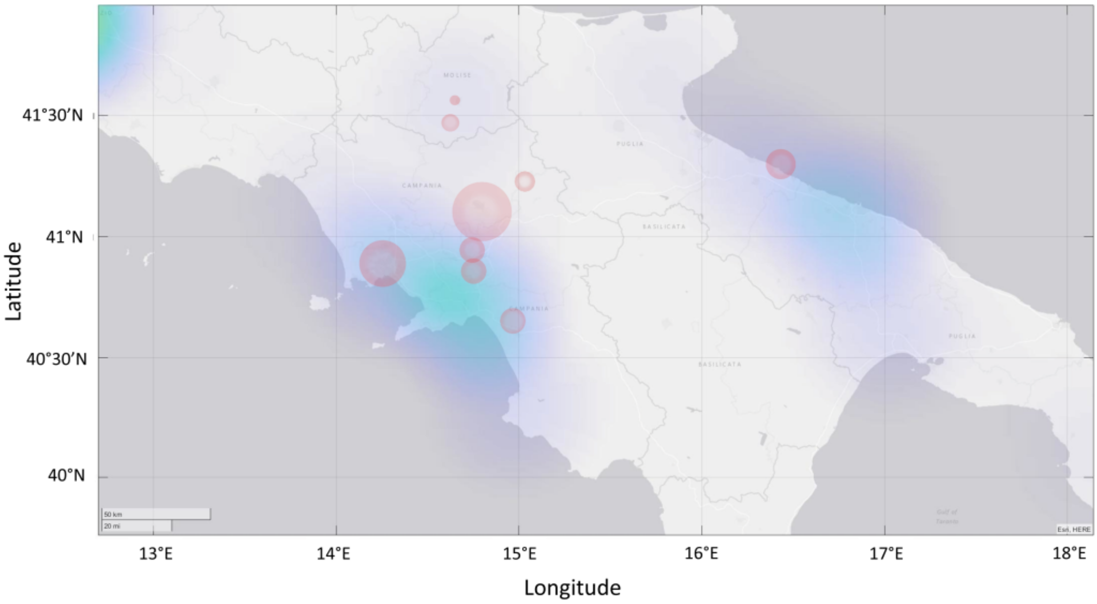

Number of close-proximity contacts (≤ 1m)
>500
<50
Latitude
45°N
40°N
30°N
Longitude
5°W
0°
5°E
10°E
15°E
20°E
25°E
30°E
35°E
FRANCE
ITALY
SPAIN
TURKEY
200 km
200 mi
Esri, HERE

Figure 5. GPS test results. Green areas indicate locations that had a low risk of SARS-CoV-2 infection. Red areas indicate locations that had a high risk of infection. The red areas correspond to locations where SARS-CoV-2 outbreaks happened during the lockdown period.


Latitude
41°30′N
41°N
40°30′N
40°N
Longitude
13°E
14°E
15°E
16°E
17°E
18°E
MOLISE
PUGLIA
CAMPANIA
BASILICATA
50 km
20 mi
Esri, HERE

Conclusions

The high correlation between CT data and the number of recorded SARS-COV-2–positive cases (with a delay of 5-7 days) was remarkable. The number of registered contacts and the number of new SARS-COV-2–positive cases showed the same weekly trend fluctuations, which not only depended on the number of tests but also on the different mobility abilities of people. Moreover, there was a time lag between the two factors, and this was the result of the incubation time of SARS-CoV-2. This time lag can be used to estimate the real incubation time of SARS-CoV-2. Further, this correlation can be extremely useful for defining and predicting infection trends and can be used to improve predictive models that only use health authorities' data. Regardless of the effectiveness of CT, the collected data provided a powerful tool for improving predictive and epidemiological models and could be integrated into different types of analyses to improve the accuracy and efficiency of predictions based on real data.

This study lays the foundation for our upcoming papers. In future papers, we will show how CT data were implemented in a CT simulator to turn it into a real data-based contagion spread simulator, which provided us with data on the mobility of the different clusters that were defined in this study. The agents' mobility data will be used to determine the risk of infection, identify epidemiological parameters, and simulate the spread of SARS-CoV-2 in different contexts. The SM-COVID-19 data set is open and free for use by the scientific community. This paper does not represent a policy pronouncement, as this would not be a scientific objective. We believe that our study may prompt informed discussions of the possible risks and likely benefits of our approach to using CT data. For these reasons, all collected data are available for further analysis.

Acknowledgments

The authors thank the Campania Region for supporting our scientific research and the development of the app. We express our gratitude to Cesare Pianese and Luca Canepa for the numerous opportunities to discuss ethical and privacy issues. We also thank the entire SM-COVID-19 team [33] for supporting and distributing the data used in this work.

Conflicts of Interest

SP and LDB are members of the academic spin-off company SM, and they were involved in the development of the SM-COVID-19 app.

Multimedia Appendix 1
Acquisition of SM-COVID-19 app data on contacts.
[DOCX File , 41 KB - jmir_v23i8e28947_app1.docx]

Multimedia Appendix 2
Open data format.
[DOCX File , 15 KB - jmir_v23i8e28947_app2.docx]

Multimedia Appendix 3
Dumping and anonymization.
[DOCX File , 15 KB - jmir_v23i8e28947_app3.docx]

Multimedia Appendix 4
Statistical analysis and estimates of the real number of SARS-CoV-2–positive cases.
[DOCX File , 14 KB - jmir_v23i8e28947_app4.docx]

Multimedia Appendix 5
Contact index and alpha values.
[DOCX File , 20 KB - jmir_v23i8e28947_app5.docx]

References

1. Coronavirus disease (COVID-19) pandemic. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 [accessed 2021-07-15]
2. Weekly epidemiological update - 5 October 2020. World Health Organization. 2020 Oct 05. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20201005-weekly-epi-update-8.pdf [accessed 2021-07-15]
3. 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 2021-07-15]
4. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. Ann Intern Med 2020 May 05;172(9):577-582 [FREE Full text] [doi: 10.7326/M20-0504] [Medline: 32150748]
5. Cho SY, Kang JM, Ha YE, Park GE, Lee JY, Ko JH, et al. MERS-CoV outbreak following a single patient exposure in an emergency room in South Korea: an epidemiological outbreak study. Lancet 2016 Sep 03;388(10048):994-1001 [FREE Full text] [doi: 10.1016/S0140-6736(16)30623-7] [Medline: 27402381]
6. Kim KH, Tandi TE, Choi JW, Moon JM, Kim MS. Middle East respiratory syndrome coronavirus (MERS-CoV) outbreak in South Korea, 2015: epidemiology, characteristics and public health implications. J Hosp Infect 2017 Feb;95(2):207-213 [FREE Full text] [doi: 10.1016/j.jhin.2016.10.008] [Medline: 28153558]

7. Shen Z, Ning F, Zhou W, He X, Lin C, Chin DP, et al. Superspreading SARS events, Beijing, 2003. Emerg Infect Dis 2004 Feb;10(2):256-260 [FREE Full text] [doi: 10.3201/eid1002.030732] [Medline: 15030693]
8. Wang SX, Li YM, Sun BC, Zhang SW, Zhao WH, Wei MT, et al. The SARS outbreak in a general hospital in Tianjin, China -- the case of super-spreader. Epidemiol Infect 2006 Aug;134(4):786-791. [doi: 10.1017/S095026880500556X] [Medline: 16371174]
9. Zhao S, Lin Q, Ran J, Musa SS, Yang G, Wang W, et al. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. Int J Infect Dis 2020 Mar;92:214-217 [FREE Full text] [doi: 10.1016/j.ijid.2020.01.050] [Medline: 32007643]
10. Domingo E. Chapter 7 - Long-term virus evolution in nature. In: Virus as Populations (Second Edition). Cambridge, Massachusetts, United States: Academic Press; 2020:225-261.
11. Fraser C, Donnelly CA, Cauchemez S, Hanage WP, Van Kerkhove MD, Hollingsworth TD, WHO Rapid Pandemic Assessment Collaboration. Pandemic potential of a strain of influenza A (H1N1): early findings. Science 2009 Jun 19;324(5934):1557-1561 [FREE Full text] [doi: 10.1126/science.1176062] [Medline: 19433588]
12. Wallinga J, Teunis P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. Am J Epidemiol 2004 Sep 15;160(6):509-516 [FREE Full text] [doi: 10.1093/aje/kwh255] [Medline: 15353409]
13. Zhang S, Diao M, Yu W, Pei L, Lin Z, Chen D. Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: A data-driven analysis. Int J Infect Dis 2020 Apr;93:201-204 [FREE Full text] [doi: 10.1016/j.ijid.2020.02.033] [Medline: 32097725]
14. Ferguson NM, Laydon D, Nedjati-Gilani G, Imai N, Ainslie K, Baguelin M, et al. Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. Imperial College COVID-19 Response Team. 2020 Mar 16. URL: https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-NPI-modelling-16-03-2020.pdf [accessed 2021-07-15]
15. Fraser C, Riley S, Anderson RM, Ferguson NM. Factors that make an infectious disease outbreak controllable. Proc Natl Acad Sci U S A 2004 Apr 20;101(16):6146-6151 [FREE Full text] [doi: 10.1073/pnas.0307506101] [Medline: 15071187]
16. Hellewell J, Abbott S, Gimma A, Bosse NI, Jarvis CI, Russell TW, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. Lancet Glob Health 2020 Apr;8(4):e488-e496 [FREE Full text] [doi: 10.1016/S2214-109X(20)30074-7] [Medline: 32119825]
17. Klinkenberg D, Fraser C, Heesterbeek H. The effectiveness of contact tracing in emerging epidemics. PLoS One 2006 Dec 20;1(1):e12. [doi: 10.1371/journal.pone.0000012] [Medline: 17183638]
18. Kretzschmar ME, Rozhnova G, Bootsma MCJ, van Boven M, van de Wijgert JHHM, Bonten MJM. Impact of delays on effectiveness of contact tracing strategies for COVID-19: a modelling study. Lancet Public Health 2020 Aug;5(8):e452-e459 [FREE Full text] [doi: 10.1016/S2468-2667(20)30157-2] [Medline: 32682487]
19. 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]
20. Bengio Y, Janda R, Yu YW, Ippolito D, Jarvie M, Pilat D, et al. The need for privacy with public digital contact tracing during the COVID-19 pandemic. Lancet Digit Health 2020 Jul;2(7):e342-e344 [FREE Full text] [doi: 10.1016/S2589-7500(20)30133-3] [Medline: 32835192]
21. Dar AB, Lone AH, Zahoor S, Khan AA, Naaz R. Applicability of mobile contact tracing in fighting pandemic (COVID-19): Issues, challenges and solutions. Comput Sci Rev 2020 Nov;38:100307 [FREE Full text] [doi: 10.1016/j.cosrev.2020.100307] [Medline: 32989380]
22. Chan J, Foster D, Gollakota S, Horvitz E, Jaeger J, Kakade S, et al. PACT: Privacy Sensitive Protocols and Mechanisms for Mobile Contact Tracing. arXiv. 2021 Preprint posted online on May 7, 2020. [FREE Full text]
23. Ahmed N, Michelin RA, Xue W, Ruj S, Malaney R, Kanhere SS, et al. A survey of COVID-19 contact tracing apps. IEEE Access 2020 Jul 20;8:134577-134601 [FREE Full text] [doi: 10.1109/access.2020.3010226]
24. Braithwaite I, Callender T, Bullock M, Aldridge RW. Automated and partly automated contact tracing: a systematic review to inform the control of COVID-19. Lancet Digit Health 2020 Nov;2(11):e607-e621 [FREE Full text] [doi: 10.1016/S2589-7500(20)30184-9] [Medline: 32839755]
25. SM-Covid-19 App — Tracing app against Covid-19. SM-COVID-19. URL: https://www.smcovid19.org [accessed 2021-07-15]
26. COVID-19 Italia - Monitoraggio situzaione. GitHub. URL: https://github.com/pcm-dpc/COVID-19/ [accessed 2021-07-15]
27. SM-COVID-19: Contact Tracing Open Data. SM-COVID-19. URL: https://smp.softmining.it/smp-smc19od.aspx [accessed 2021-07-15]
28. Yang L, Dai J, Zhao J, Wang Y, Deng P, Wang J. Estimation of incubation period and serial interval of COVID-19: analysis of 178 cases and 131 transmission chains in Hubei province, China. Epidemiol Infect 2020 Jun 19;148:e117 [FREE Full text] [doi: 10.1017/S0950268820001338] [Medline: 32594928]
29. Xiao Y, Torok ME. Taking the right measures to control COVID-19. Lancet Infect Dis 2020 May;20(5):523-524 [FREE Full text] [doi: 10.1016/S1473-3099(20)30152-3] [Medline: 32145766]

30. Aleta A, Martín-Corral D, Piontti APY, Ajelli M, Litvinova M, Chinazzi M, et al. Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. Nat Hum Behav 2020 Sep;4(9):964-971. [doi: 10.1038/s41562-020-0931-9] [Medline: 32759985]
31. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). Science 2020 May 01;368(6490):489-493 [FREE Full text] [doi: 10.1126/science.abb3221] [Medline: 32179701]
32. Galvani AP, May RM. Epidemiology: dimensions of superspreading. Nature 2005 Nov 17;438(7066):293-295 [FREE Full text] [doi: 10.1038/438293a] [Medline: 16292292]
33. Team — SM-Covid-19 App. SM-COVID-19. URL: https://www.smcovid19.org/team/ [accessed 2021-07-20]

Abbreviations

CI: contact index
CT: contact tracing
MERS-COV: Middle East respiratory syndrome coronavirus
SARS-CoV: severe acute respiratory syndrome coronavirus
SM: SoftMining
SSE: superspreading event

Edited by C Basch; submitted 19.03.21; peer-reviewed by F Tonelli, A Lahiri; comments to author 29.04.21; revised version received 10.05.21; accepted 15.05.21; published 02.08.21.

Please cite as:
Piotto S, Di Biasi L, Marrafino F, Concilio S
Evaluating Epidemiological Risk by Using Open Contact Tracing Data: Correlational Study
J Med Internet Res 2021;23(8):e28947
URL: https://www.jmir.org/2021/8/e28947
doi:10.2196/28947
PMID:34227997

©Stefano Piotto, Luigi Di Biasi, Francesco Marrafino, Simona Concilio. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 02.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Association Between Public Opinion and Malaysian Government Communication Strategies About the COVID-19 Crisis: Content Analysis of Image Repair Strategies in Social Media

Nasaai Masngut[1*], MSocSci; Emma Mohamad[2,3*], PhD

[1]Malaysian Nuclear Agency, Bangi, Malaysia

[2]Centre for Research in Media and Communication, Faculty of Social Sciences and Humanities, Universiti Kebangsaan Malaysia, Bangi, Malaysia

[3]UKM x UNICEF Communication for Development Centre in Health, Faculty of Social Sciences and Humanities, Universiti Kebangsaan Malaysia, Bangi, Malaysia

[*]all authors contributed equally

Corresponding Author:
Emma Mohamad, PhD
Centre for Research in Media and Communication
Faculty of Social Sciences and Humanities
Universiti Kebangsaan Malaysia
Jalan Nik Ahmed Kamil
Bangi, 43600
Malaysia
Phone: 60 389215456
Email: emmamohamad@ukm.edu.my

Abstract

Background: The COVID-19 health crisis has posed an unprecedented challenge for governments worldwide to manage and communicate about the pandemic effectively, while maintaining public trust. Good leadership image in times of a health emergency is paramount to ensure public confidence in governments' abilities to manage the crisis.

Objective: The aim of this study was to identify types of image repair strategies utilized by the Malaysian government in their communication about COVID-19 in the media and analyze public responses to these messages on social media.

Methods: Content analysis was employed to analyze 120 media statements and 382 comments retrieved from Facebook pages of 2 mainstream newspapers—Berita Harian and The Star. These media statements and comments were collected within a span of 6 weeks prior to and during the first implementation of Movement Control Order by the Malaysian Government. The media statements were analyzed according to Image Repair Theory to categorize strategies employed in government communications related to COVID-19 crisis. Public opinion responses were measured using modified lexicon-based sentiment analysis to categorize positive, negative, and neutral statements.

Results: The Malaysian government employed all 5 Image Repair Theory strategies in their communications in both newspapers. The strategy most utilized was reducing offensiveness (75/120, 62.5%), followed by corrective action (30/120, 25.0%), evading responsibilities (10/120, 8.3%), denial (4/120, 3.3%), and mortification (1/120, 0.8%). This study also found multiple substrategies in government media statements including denial, shifting blame, provocation, defeasibility, accident, good intention, bolstering, minimization, differentiation, transcendence, attacking accuser, resolve problem, prevent recurrence, admit wrongdoing, and apologize. This study also found that 64.7% of public opinion was positive in response to media statements made by the Malaysian government and also revealed a significant positive association (P=.04) between image repair strategies utilized by the Malaysian government and public opinion.

Conclusions: Communication in the media may assist the government in fostering positive support from the public. Suitable image repair strategies could garner positive public responses and help build trust in times of crisis.

(J Med Internet Res 2021;23(8):e28074) doi:10.2196/28074

KEYWORDS

COVID-19; crisis; health communication; image repair; Malaysian government; sentiment; communication; content analysis; public opinion; social media; strategy

Introduction

Background

A public health crisis is often a threat to health infrastructure because it can cripple an existing health care system on a larger scale. As such, a global pandemic such as the COVID-19 pandemic, could potentially be damaging to a national health system [1,2] and the stability of a country. Much of the success of a public health campaign is determined by how far the public trust their government and health institution. During the 2018 Ebola outbreak in the Democratic Republic of Congo, the federal government faced a critical challenge in disseminating relevant information, obtaining local cooperation, and conducting mass inoculation programs. This was due to a weak social system and diminished public trust. The civil unrest and domestic terrorism in the Democratic Republic of Congo made it more difficult for the government to combat the Ebola virus [3].

In Malaysia, the sudden outbreak of severe acute respiratory syndrome (SARS) in 2003 initially proved too much for the government to handle. Constant communication on SARS in the mass media led to a mild panic among the public [4]. However, the Ministry of Health played a vital role in mitigating the situation by providing constant situational reports and mobilizing a crisis management team. The Ministry also solidified their role as the main source of information to the public through daily updates on SARS and active public engagements [4]. Communication from health authorities is important to ensure that the public take necessary actions to control the spread of infectious diseases [5,6] by framing issues that are deemed critical for public's safety [7].

While keeping the public informed remains the main goal of crisis communication, earning public trust is another matter. Gaining public trust is often achieved by using credible means [8], which can effectively improve message reception. However, social media has become an important source of information regardless of its authenticity and reliability [9]. The spread of misinformation has proved to be dangerous in times of crisis because it can easily undermine health efforts and sow the seeds of doubt among the public. Earning trust is vital to allow the government to implement necessary measures and obtain cooperation from the public.

COVID-19 was first detected in Malaysia in January 2020. Since then, the Malaysian Government through the Ministry of Health has been consistent in delivering daily updates on COVID-19 to the public. As part of their integrated effort in combatting COVID-19, the Malaysian government has activated its Crisis Preparedness and Response Center, which is a part of an overall strategy to overcome the pandemic. This center acted as a central command for early outbreak coordination and management control over the pandemic [10].

Even though the Malaysian government was generally commended for their swift action during the initial stage of local COVID-19 transmission, there were several challenges to secure public's trust at the time. This was largely due to an unexpected political upheaval after the Prime Minister's resignation in early March 2020. The sudden government transition itself puts Malaysia in a fragile state because of uncertainty about government stability and the capability of the newly formed government to combat COVID-19 [11]. In addition, a premature statement made by the Health Minister, suggesting that warm water might kill COVID-19 virus, was ridiculed, which did not help to foster public trust [12]. This controversial statement was later criticized by many medical practitioners who refuted the claim. In another occasion, the minister was teased for his lack of general knowledge when he accidentally claimed that 500 countries had participated in a recent World Health Organization video conference [13]. Controversial statements such as these could damage government's reputation and jeopardize public's trust in the government's ability to manage the pandemic. Furthermore, several statements from the government were found to be contradictory [14], which led to escalating confusion among the public. Even though providing contradictory information in early phases of a crisis is not uncommon [15], such inconsistency may lead to further distrust among the public.

Therefore, we examined how the Malaysian government managed crisis communications on COVID-19 in the media. Specifically, this study identifies image repair strategies utilized by the Malaysian government in the media and examines how society react to these strategies by observing the directions of public sentiments through social media responses. Findings from this research could be useful in planning better crisis communication strategies.

Measures to Mitigate COVID-19

Daily updates on COVID-19 were given by the Director General of Health, while the Minister of Defense was in charge of communicating standard operating procedures and public measures to mitigate the spread of the virus. These updates were given through separate press conferences and published in different forms of mass media, including Telegram, a freeware instant messaging app that could reach millions of people rapidly. In addition, to ensure information on COVID-19 was disseminated through proper channels, a dedicated website was set up under the purview of Ministry of Health [16]. This website is a centralized platform for daily updates on cases as well as any new developments on COVID-19, in addition to the daily press conferences. To step up their effort in curbing COVID-19, the Malaysian government enacted a Movement Control Order on March 18, 2020 for a period of 2 weeks, then extended the order by 5 weeks (for a total 7 weeks). The Movement Control Order was executed under the purview of the National Security Council under the Infectious Disease Control and Prevention Act 1988, as well as the Police Act 1967 [17]. This special measure allowed the government to enforce various movement restrictions on general population, in order to reduce virus transmission in the population.

Key economic areas and activities came to halt, along with education sector, religious houses, and sports and cultural activities. However, movement restriction exceptions were given for essential services. The implementation of Movement Control Order later transitioned to a Conditional Movement Control Order, which eased economic restriction while maintaining movement restriction. When the local case transmission showed improvement, the government introduced a Recovery Movement Control Order, which allowed gradual reopening of most key sectors. Systematic implementation of the different levels of Movement Control Order demonstrated the Malaysian government's commitment in tackling COVID-19 nationwide. The Malaysian government also implemented MySejahtera, a contact-tracing system mobile device app. MySejahtera provided a database of COVID-19 cases in communities and charts of patients' prior movements, to help warn the public about hotspots [18]. Some 6.2 million users downloaded the MySejahtera app by July 2020, indicating a high-level voluntary usage of the tracking app among the public [19]. The app also received an award at the Ministry of Health Innovation Day 2020 that recognized its robust features and contribution in tackling the pandemic [20]

Image Repair Framework

This study utilized Image Repair Theory [21] to analyze government communications in the media. The theory proposed 5 main strategies and 15 substrategies of image repair. The main strategies proposed by the Image Repair Theory are denial, evasion of responsibility, reducing offensiveness, corrective action, and mortification. These strategies can be utilized independently or collectively to improve public perception. Image repair strategies are frequently used by organizations in crisis to gain favorable responses from the public. Studies [22-24] have also showed how different image repair strategies were able to help improve government's image in crisis situations.

One study [25] suggested that a majority of the public form their perception of risk of a public health crisis from the media they consume. Various media frames may be used to deliver public health messages to help contextualize the message and urge the public to take action. Among the commonly utilized frames are risk magnitude, self-efficacy, episodic framing and economic uncertainty, which are used to communicate symptoms, likeliness to contract a disease, or protective measures the public may undertake [7]. Restoring damaged perceptions may require effective framing contexts to deliver key messages to the public. This is because the media often use sensationalism to ensure engagement from the audience [8] and this may jeopardize individual's or organization's image and reputation. It has also been suggested that studying the trend of public attention in the media may also help authorities determine appropriate frames to deliver key messages effectively [26]. Even though this study did not specifically investigate framing in the media, it is important to note that media messages can play a huge role in influencing public behavior [27].

While the theory [21] categorizes image repair strategies, it does not propose that one strategy is better than the other. However, it is useful to look at how these strategies are utilized and how the public responds to them in different context. For example, a study [22] found that despite different strategies used to improve Chinese government's image during the SARS crisis, the effort was unsuccessful due to frequent contradictory messages. In another study [23], the Chinese government was successful at restoring public's confidence using denial, bolstering and corrective action strategies in their communications. Similar strategies were also utilized by the Saudi Arabia government when faced by accusations of terrorism back in 2003 [24].

Although prior studies have recognized the importance of image repair in various crises, investigation of public sentiment on image repair strategies has been limited. Therefore, to address this knowledge gap, we explored the following research questions: (1) What are the image repair strategies employed by the Malaysian government in their communications on COVID-19 in the media? To investigate which image repair strategies utilized in the media by the Malaysian government on COVID-19, it is pertinent to analyze media statements by government officials. These statements may contain either one or more strategies outlined in Benoit's image repair theory. (2) What is the public sentiment towards COVID-19 media statements from the Malaysian government? Government communications on COVID-19 may bring about different kinds of sentiments from the public. As the direction of sentiment may become an initial indicator of the effectiveness of image repair efforts, this study will scrutinize public's responses toward COVID-19 media statements by the Malaysian government. (3) What is the relationship between image repair strategies and the direction of public sentiment on COVID-19 media statements by the Malaysian government? By examining the relationship between image repair strategies and the direction of public sentiment, this study aims to describe the effectiveness of the Malaysian government's effort in communicating about COVID-19.

Methods

Data Collection

This study selected 2 mainstream newspapers—Berita Harian and The Star—for data collection. Statements made by the Malaysian government were retrieved via online Facebook platforms (Berita Harian Online and The Star Facebook page). These 2 newspapers have a broad influence, with 5.7 million (Berita Harian) and 1.2 million (The Star) page followers in Malaysia. Due to their prominence and influential nature, media statements from selected officials (ie, the Prime Minister, Ministers, Director General of Health and Inspector General of Police) were chosen. Additionally, responses to questions from several ministries were also included. Online comments from the public on statements made by these government officials were also collected to answer the second research question.

Sample Period

Key official media statements published in Berita Harian Online and The Star Facebook page pertaining to COVID-19 from March 4, 2020 (2 weeks before the implementation of Movement Control Order) to April 30, 2020 (6 weeks into the implementation of the Movement Control Order) were collected.

XSL•FO RenderX

Online comments from the statements were also extracted from the same period.

Sample Size and Sample Data Collection

To investigate the first research question, 120 media statements were taken as a sample. Subsequently, we estimated [28] a sample size of 382 out of 59,941 online comments to examine the direction of public sentiment; this sample has a confidence level of 95% and margin of error of 5%. sampling with stratification (50% from Berita Harian Online and 50% from The Star Facebook page), by selecting top comments labeled as “most relevant” by Facebook algorithm from each media statement, was used. The algorithm sorted and ranked comments based on the highest quality of comments that are relevant to users [29]. Only comments with Malaysian names were taken as sample [30]. These comments were then included for content analysis based on image repair framework.

Content Analysis and Coding Scheme

Content analysis was employed to identify image repair strategies utilized in the 120 online media statements by the Malaysian government and 382 comments from the public. Coding scheme for research question 1 was adapted from a previous study [31], in which 5 main strategies and 15 substrategies were used. The strategies are denial, evading responsibility, reducing offensiveness, corrective action, and mortification. The substrategies were simple denial, shifting blame, provocation, defeasibility, accident, good intention, bolstering, minimization, differentiation, transcendence, attacking accuser, fixing problem, preventing recurrence, admitting fault, and apologizing.

To answer research question 2, this study categorized sentiments by employing Valence Aware Dictionary for Sentiment Reasoning (VADER) text analysis technique [32]. Through this particular sentiment analysis method, public comments were categorized into positive, negative or neutral. Despite the language limitation of the VADER technique, which only catered to English language [33], this study was able to identify and replicate its algorithm to the Malay language for when analyzing comments in Bahasa Malaysia. To correspond with VADER analytical algorithm, 4 main principles were applied toward analyzing social media text in online comments: punctuation, capitalization, intensifiers, and conjunction [33]. Results obtained from research question 1 and research question 2 were then to measure relationships outlined in research question 3.

Validity and Intercoder Reliability

Validity for coding instrument in content analysis can be described as face validity, construct validity, content validity, and criteria validity [34]. The coding instrument for the study has been face validated by a researcher based at the *Universiti Kebangsaan Malaysia*.

Two coders analyzed the media statements and public comments. To ensure intercoder reliability, both coders were trained using a coding scheme. Intercoder reliability [35] was calculated as $R = 2M / (N1 + N2)$, where M is the total mutually agreed coding result and $N1$ and $N2$ are numbers of decisions made by coder 1 and coder 2. Each coder analyzed 10 media statements that were not part of the sample (R=0.74) and 10 public comments (R=0.74). An R value equal or more than 0.70 is accepted as reliable [36].

Statistical Analysis

A statistical analysis was performed using SPSS software (version 22; IBM Corp). For descriptive analysis, frequency tables on media source, publication phase, statement source and image repair strategies (including substrategies) were created to address research question 1. This analysis revealed highest frequency of image repair strategies and substrategies utilized by the Malaysian government. A frequency table was also generated to categorize and reveal direction of public sentiments (positive, negative or neutral) to answer research question 2. Chi-square tests were used to determine the association between COVID-19 image repair strategies by the Malaysian government and the direction of public sentiment from online comments to answer research question 3.

To measure the effect size for cross-tabulated data, Pallant [37] has suggested that for a table larger than 2 by 2 Cramer V, which takes into account degrees of freedom, with V<0.07 small, V=0.08-0.21 medium, and V>0.35 large suggested.

Results

The *reducing offensiveness* strategy was the most utilized strategy in COVID-19 media statements (75/120, 62.5%). It was used more than the *corrective action* strategy, which was the second-most used (30/120, 25.0%) (Table 1).

For image repair substrategies, a total of 454 statements were coded (Multimedia Appendix 1). The *bolstering positive quality* substrategy (reducing offensiveness strategy) appeared the most in the media statements (108/454, 23.7%), while the *fixing problem* substrategy (under corrective action strategy) came in second (104/454, 22.9%). Other frequently occurring substrategies were preventing recurrence (102/454, 22.4%) and minimization (38/454, 5.3%).

Table 1. Image repair strategy frequency.

Image repair strategy	Instances, n (%)
Denial	10 (8.3)
Evade responsibility	4 (3.3)
Reduce offensiveness	75 (62.5)
Corrective action	30 (25.0)
Mortification	1 (0.8)
Total	120 (100.0)

Findings (Multimedia Appendix 2) also revealed that most media statements were made by the Director General of Health (24/120, 20.0%), followed closely by the Prime Minister and Senior Minister of Defense, (both 22/120, 18.3%). In terms of media sources, Berita Harian published more media statements from the Government on COVID-19 (87/120, 72.5%) than The Star (33/120, 27.5%). This study also observed that most media statements were published during the Movement Control Order (105/120, 87.5%) as opposed to before the Movement Control Order (15/120, 12.5%) (Multimedia Appendix 3).

A sample of Facebook comments (382/59,941) from the 120 media statements showed more positive (247/382, 64.7%), than negative (86/382, 22.5%) or neutral (49/382, 12.8%) sentiments. Berita Harian had slightly more positive sentiments (134/382, 70.9%) than The Star (112/382, 58.3%). Positive sentiments were higher in both media; however, there were more negative sentiments (Berita Harian: 34/382, 18.0%; The Star: 52/382, 27.1%) than neutral sentiments (Berita Harian: 22/382, 11.1%; The Star: 27/382, 14.6%) recorded.

The *denial* strategy received the most positive sentiments (33/47, 70.2%) (Table 2), while the *corrective action* strategy came second (42/60, 70.0%). The *reducing offensiveness* strategy, which was most utilized in Government statements, was the third highest in receiving positive sentiments (169/264, 64.0%). There was a significant association between image repair strategies of the Malaysian government COVID-19 media statements and the direction of public sentiment with a small effect (n=382, χ_8^2= 0.146, *P*=0.039 Cramer *V*=0.039).

Table 2. Direction of sentiment.

Direction of sentiment	Strategy, n (%)				
	Denial (n=47)	Evade responsibility (n=10)	Reduce offensiveness (n=264)	Corrective action (n=60)	Mortification (n=1)
Positive	33 (70.2)	2 (20.0)	169 (64.0)	42 (70.0)	1 (100.0)
Negative	6 (12.8)	4 (40.0)	66 (25.0)	10 (16.7)	0 (0)
Neutral	8 (17.0)	4 (40.0)	29 (11.0)	8 (13.3)	0 (0)

Discussion

Principal Results

A range of image repair strategies were employed by the Malaysian government in their COVID-19 media statements and could be categorized according to the 5 strategies suggested by Image Repair Theory [21]. In particular, the *reduce offensiveness* strategy was the most utilized strategy in communicating COVID-19 in the media, and the majority of these media statements employed the *bolstering positive quality* substrategy. There was also a significant association (*P*=.04) between image repair strategies and the direction of public sentiment. Although the *reducing offensiveness* strategy was most utilized, results showed that the *denial* strategy received the highest positive sentiments, which was followed by the *corrective action* strategy, and then the *reducing offensiveness* strategy. Nonetheless, overall sentiment towards government's messaging were positive.

Although the image repair theory does not stipulate crisis management strategies employed by the Malaysian government, the theory provided a framework to analyze their statements in the media [38]. An understanding of how a crisis is framed by the media helps to identify strategies that might work in future crisis communications.

This study revealed that the strategy utilized most by the Malaysian government in their COVID-19 media statements was reducing offensiveness. An example of reducing offensiveness was a statement by the Health Minister explaining the need to minimize visiting hours at the hospitals as a proactive measure to reduce the risk of COVID-19 infection. A previous study [39] revealed a similar result; reducing offensiveness in the wake of Russian tourism ban to Egypt due to safety and security concern was proven to be effective and has resulted in Russian flight resuming its operations to Cairo in April 2018 after a 2-year ban.

Implications, Limitations, and Future Work

This study contributes to current literature on COVID-19 in Malaysia [40,41] as well as understanding COVID-19 crisis communications by the Malaysian government based on the framework suggested by Benoit [21]. It adds to the body of knowledge on image repair strategies and public opinion, which may be useful in a health crisis with this global magnitude.

Government efforts to mitigate the spread of COVID-19, such as enacting the Movement Control Order, have had a big impact on population, such as closure of school and non-essential businesses. Therefore, communication of these types of orders must be done correctly to not instigate fear and panic among the public [42]. Negative sentiment from the public could hamper government efforts and cause distrust in the health care delivery system.

A prior study found that, in Malaysia, television and internet news portals are primarily used to access information on COVID-19 [40] and suggested that health authorities should pay considerable attention to the use of appropriate media channels and sources to allow for more effective dissemination of critical information to the public. By identifying too which image repair strategies the Malaysian public responded well, the findings of our study provides insight into information framing that can receive positive responses from the public. We suggest using the denial, corrective action, and reducing offensiveness in television and internet news portals to communicate about crises.

The government should evaluate strengths and limitations of a country in addressing a health crisis [43]. A previous study [44] highlighted several challenges in communicating in a crisis including misinformation, lack of guidance, and information leakage. It has also been suggested that social media caused more confusion, rather than consolidating public effort against the pandemic [45]. Therefore, information in the media must focus on improving trust, building social solidarity, and reducing chaos, while educating the public on prevention measures and reducing burden on the health system. With suitable and effective image repair strategies, the government could minimize public uncertainty and mitigate the spread of false information.

One of the limitations of this study pertains to the sampling period. As the media statements and comments were captured in a specific time span, the results obtained are only applicable and true to the specific time range. Additionally, the study only took public opinions from Facebook comments into account. Results may have differed if the study extrapolated samples in other social media such as Twitter or Instagram. In addition, the consistent appearance of well-liked individuals such as the Director General of Health, may have contributed to the overall positive responses. The Director General of Health was considered a national hero and an exemplary leader in crisis, and his appearances in front of the camera to deliver daily updates on COVID-19 in a calm and composed manner has earned him the image of a rationale leader, providing assurance and tranquility to the public [46]. An investigation into the roles of frequent media appearances and leadership figures in times of crisis, as well as its influences toward public acceptance should be explored in future studies.

Conclusions

This study provided comprehensive insight into image repair strategies in the media by the Malaysian government and how members of the public reacted in response to these strategies. The findings of this study could be useful to advise future crisis communication planning, particularly in a health crisis.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Image repair substrategy frequency distribution.
[DOCX File , 13 KB - jmir_v23i8e28074_app1.docx]

Multimedia Appendix 2
Source frequency distribution.
[DOCX File , 13 KB - jmir_v23i8e28074_app2.docx]

Multimedia Appendix 3
Publication phase distribution.
[DOCX File , 13 KB - jmir_v23i8e28074_app3.docx]

References

1. Burkle FM. Challenges of global public health emergencies: development of a health-crisis management framework. Tohoku J Exp Med 2019 Sep;249(1):33-41 [FREE Full text] [doi: 10.1620/tjem.249.33] [Medline: 31548492]
2. Al-Jazairi A. Disasters and disaster medicine. In: Essentials of Accident and Emergency Medicine. London UK: Intech Open; 2017:93-118.
3. Vinck P, Pham PN, Bindu KK, Bedford J, Nilles EJ. Institutional trust and misinformation in the response to the 2018-19 Ebola outbreak in North Kivu, DR Congo: a population-based survey. Lancet Infect Dis 2019 May;19(5):529-536. [doi: 10.1016/S1473-3099(19)30063-5] [Medline: 30928435]
4. Sanusi S, Siarap K. Peranan Perhubungan Awam dan Komunikasi Krisis: Kajian Terhadap Kementerian Kesihatan Malaysia Dalam Penanganan Wabak SARS. E-Bangi Journal of Social Science and Humanities 2014;9(2):125-134 [FREE Full text]

5. Ngai CSB, Singh RG, Lu W, Koon AC. Grappling with the COVID-19 health crisis: content analysis of communication strategies and their effects on public engagement on social media. J Med Internet Res 2020 Aug 24;22(8):e21360 [FREE Full text] [doi: 10.2196/21360] [Medline: 32750013]
6. Odlum M, Yoon S. What can we learn about the Ebola outbreak from tweets? Am J Infect Control 2015 Jun;43(6):563-571. [doi: 10.1016/j.ajic.2015.02.023] [Medline: 26042846]
7. Sandell T, Sebar B, Harris N. Framing risk: communication messages in the Australian and Swedish print media surrounding the 2009 H1N1 pandemic. Scand J Public Health 2013 Dec 19;41(8):860-865. [doi: 10.1177/1403494813498158] [Medline: 23873631]
8. Kott A, Limaye RJ. Delivering risk information in a dynamic information environment: framing and authoritative voice in Centers for Disease Control (CDC) and primetime broadcast news media communications during the 2014 Ebola outbreak. Soc Sci Med 2016 Nov;169:42-49 [FREE Full text] [doi: 10.1016/j.socscimed.2016.09.029] [Medline: 27682022]
9. Phillips HW, Chen J, Wilson B, Udawatta M, Prashant G, Nagasawa D, et al. Social media use for professional purposes in the neurosurgical community: a multi-institutional study. World Neurosurg 2019 Sep;129:e367-e374. [doi: 10.1016/j.wneu.2019.05.154] [Medline: 31132504]
10. Kementerian Kesihatan Malaysia. Crisis Preparedness and Response Center. URL: http://covid-19.moh.gov.my/ [accessed 2021-05-02]
11. Amin K. Covid-19: Keputusan Muhyiddin ambil alih kerajaan tidak kena pada masa. Sinar Harian. 2020 Mar 16. URL: https://www.sinarharian.com.my/article/74038/KHAS/Covid-19/Covid-19-Keputusan-Muhyiddin-ambil-alih-kerajaan-tidak-kena-pada-masa [accessed 2021-05-02]
12. Said S. Minum banyak air cegah virus masuk badan? Dr Imelda minta orang ramai ikut apa yang ditulis di web KKM. Portal Mstar. 2020 Mar 22. URL: https://www.mstar.com.my/lokal/viral/2020/03/22/dr-imelda-air-suam [accessed 2021-05-02]
13. Zainal H. Adham Baba: I misspoke, I meant 500 participants. The Star. 2020 Apr 19. URL: https://www.thestar.com.my/news/nation/2020/04/19/adham-baba-i-misspoke-i-meant-500-participants [accessed 2021-05-02]
14. Ong BT. Pengumuman Menteri mengelirukan. Sinar Harian. URL: https://www.sinarharian.com.my/article/106842/KOLUMNIS/Pengumuman-menteri-mengelirukan [accessed 2020-10-27]
15. Avery EJ, Kim S. Anticipating or precipitating crisis? health agencies may not be heeding best practice advice in avian flu press releases. Public Relat Rev 2009 Apr 14;21(2):187-197. [doi: 10.1080/10627260802557449]
16. Situasi Semasa Pandemik COVID-19 Di Malaysia. Covid-19 Malaysia. URL: http://covid-19.moh.gov.my [accessed 2020-07-10]
17. Covid-19. National Security Council Malaysia. URL: https://www.mkn.gov.my/web/ms/Covid-19/ [accessed 2020-07-10]
18. Noh MF. Kempen Pembudayaan Norma baharu: Memperkasa usaha melawan Covid-19. BH Online. URL: https://www.bharian.com.my/berita/nasional/2020/08/719110/kempen-pembudayaan-norma-baharu-memperkasa-usaha-melawan-covid-19 [accessed 2020-06-28]
19. Povera A. 2 million have downloaded MySejahtera app. New Straits Times. URL: https://www.nst.com.my/news/nation/2020/07/609112/62-million-have-downloaded-mysejahtera-app [accessed 2020-06-28]
20. Fuad F. Mysejahtera johan kategori teknologi Hari Inovasi KKM 2020. BH Online. URL: https://www.bharian.com.my/berita/nasional/2021/03/800614/mysejahtera-johan-kategori-teknologi-hari-inovasi-kkm-2020 [accessed 2020-03-27]
21. Elsbach KD, Benoit WL. Accounts, excuses, and apologies: a theory of image restoration strategies. Adm Sci Q 1997 Sep;42(3):584. [doi: 10.2307/2393739]
22. Zhang E, Benoit WL. Former minister Zhang's discourse on SARS: government's image restoration or destruction? Public Relat Rev 2009 Sep;35(3):240-246 [FREE Full text] [doi: 10.1016/j.pubrev.2009.04.004] [Medline: 32288050]
23. Peijuan C, Ting LP, Pang A. Managing a nation's image during crisis: a study of the Chinese government's image repair efforts in the “Made in China” controversy. Public Relat Rev 2009 Sep;35(3):213-218. [doi: 10.1016/j.pubrev.2009.05.015]
24. Zhang J, Benoit W. Message strategies of Saudi Arabia's image restoration campaign after 9/11. Public Relat Rev 2004:161-167 [FREE Full text] [doi: 10.1016/j.pubrev.2004.02.006]
25. Roche JP, Muskavitch MAT. Limited precision in print media communication of West Nile virus risks. Sci Commun 2016 Aug 17;24(3):353-365. [doi: 10.1177/1075547002250300]
26. Menashe CL, Siegel M. The power of a frame: an analysis of newspaper coverage of tobacco issues--United States, 1985-1996. J Health Commun 1998;3(4):307-325. [doi: 10.1080/108107398127139] [Medline: 10977260]
27. 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]
28. Krejcie R, Morgan D. Determining sample size for research activities. Educ Psychol Meas 1970;30:607-610 [FREE Full text] [doi: 10.1177/001316447003000308]
29. What does most relevant mean on a Facebook page post. FaceBoook Help Center. URL: https://www.facebook.com/help/539680519386145 [accessed 2020-06-28]
30. Top 100 Malaysian names. Students of the World. URL: http://www.studentsoftheworld.info/penpals/stats.php?Pays=MLS [accessed 2020-07-15]

31. Arandas M. Malaysian Media Coverage of Palestinian Presidents? Image During Crises 1996-2016. Malaysia: Universiti Kebangsaan Malaysia; 2019.
32. Hutto C, Gilbert E. VADER: A parsimonious rule-based model for sentiment analysis of social media text. 2015 Presented at: Eighth International AAAI Conference on Weblogs and Social Media; June 1-4; Ann Arbor, Michigan URL: https://www.researchgate.net/publication/275828927_VADER_A_Parsimonious_Rule-based_Model_for_Sentiment_Analysis_of_Social_Media_Text [doi: 10.25148/etd.fi14110776]
33. Pandey P. Simplifying sentiment analysing using VADER in Python on social media text. Analytics Vidhya. URL: https://medium.com/analytics-vidhya/simplifying-social-media-sentiment-analysis-using-vader-in-python-f9e6ec6fc52f [accessed 2020-06-28]
34. Rourke L, Anderson T. Validity in quantitative content analysis. Educ Technol Res Dev 2004 Mar;52(1):5-18. [doi: 10.1007/BF02504769]
35. Holsti O. Content Analysis for The Social Sciences and Humanities. London: Addison-Wesley; 1969.
36. Lombard M, Snyder-Duch J, Bracken CC. Practical resources for assessing and reporting intercoder reliability in content analysis research. Research Gate. 2004. URL: https://www.researchgate.net/publication/242785900_Practical_Resources_for_Assessing_and_Reporting_Intercoder_Reliability_in_Content_Analysis_Research_Projects [accessed 2021-06-02]
37. Pallant J. SPSS Survival Manual: A Step-by-Step Guide to Data Analysis Using IBM SPSS. New Jersey: Routledge; 2013.
38. An S, Gower KK. How do the news media frame crises? a content analysis of crisis news coverage. Public Relat Rev 2009 Jun;35(2):107-112. [doi: 10.1016/j.pubrev.2009.01.010]
39. Moneer A. Egypt's image repair strategies in time of crisis: a case study of the Russian flight crash. Zeitschrift für Tourismuswissenschaft 2020;12(2):206-220. [doi: 10.1515/tw-2020-0002]
40. Mohamad E, Tham JS, Ayub SH, Hamzah MR, Hashim H, Azlan AA. Relationship between COVID-19 information sources and attitudes in battling the pandemic among the Malaysian public: cross-sectional survey study. J Med Internet Res 2020 Nov 12;22(11):e23922 [FREE Full text] [doi: 10.2196/23922] [Medline: 33151897]
41. Azlan AA, Hamzah MR, Sern TJ, Ayub SH, Mohamad E. Public knowledge, attitudes and practices towards COVID-19: a cross-sectional study in Malaysia. PLoS One 2020 May 21;15(5):e0233668 [FREE Full text] [doi: 10.1371/journal.pone.0233668] [Medline: 32437434]
42. Mohamad E, Azlan A. COVID-19 and communication planning for health emergencies. Malaysian J Commun 2020:1-2 [FREE Full text]
43. Azlan AA, Hamzah MR, Tham JS, Ayub S, Ahmad AL, Mohamad E. Associations between health literacy and sociodemographic factors: a cross-sectional study in Malaysia utilising the HLS-M-Q18. Int J Environ Res Public Health 2021 May 02;18(9):4860 [FREE Full text] [doi: 10.3390/ijerph18094860] [Medline: 34063294]
44. Bao H, Cao B, Xiong Y, Tang W. Digital media's role in the COVID-19 pandemic. JMIR Mhealth Uhealth 2020 Sep 18;8(9):e20156 [FREE Full text] [doi: 10.2196/20156] [Medline: 32530817]
45. Wong FHC, Liu T, Leung DKY, Zhang AY, Au WSH, Kwok WW, et al. Consuming information related to COVID-19 on social media among older adults and its association with anxiety, social trust in information, and COVID-safe behaviors: cross-sectional telephone survey. J Med Internet Res 2021 Feb 11;23(2):e26570 [FREE Full text] [doi: 10.2196/26570] [Medline: 33523825]
46. Wissgott S. In the spotlight? the doctors at the top: truth-tellers and heartthrobs. China Global TV Network. 2020. URL: https://news.cgtn.com/news/2020-04-14/In-the-Spotlight-Doctors-at-the-top-Truth-tellers-and-heartthrobs-PGIyev8g9y/index.html [accessed 2020-06-20]

Abbreviations

SARS: severe acute respiratory syndrome
VADER: Valence Aware Dictionary for Sentiment Reasoning

Edited by C Basch; submitted 19.02.21; peer-reviewed by R Drávucz, H Moon; comments to author 21.04.21; revised version received 11.05.21; accepted 17.05.21; published 04.08.21.

Please cite as:
Masngut N, Mohamad E
Association Between Public Opinion and Malaysian Government Communication Strategies About the COVID-19 Crisis: Content Analysis of Image Repair Strategies in Social Media
J Med Internet Res 2021;23(8):e28074
URL: https://www.jmir.org/2021/8/e28074
doi:10.2196/28074
PMID:34156967

©Nasaai Masngut, Emma Mohamad. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 04.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Exploring Changes to the Actionability of COVID-19 Dashboards Over the Course of 2020 in the Canadian Context: Descriptive Assessment and Expert Appraisal Study

Erica Barbazza[1*], MSc; Damir Ivanković[1*], MBA, MD; Sophie Wang[2,3], MPH; Kendall Jamieson Gilmore[4], MSc; Mircha Poldrugovac[1], MSc, MD; Claire Willmington[4], MSc; Nicolas Larrain[2,3], MA; Véronique Bos[1], MA; Sara Allin[5], MSc, PhD; Niek Klazinga[1], MD, PhD; Dionne Kringos[1], MSc, PhD

[1]Department of Public and Occupational Health, Amsterdam UMC, Amsterdam Public Health Research Institute, University of Amsterdam, Amsterdam, Netherlands

[2]OptiMedis AG, Hamburg, Germany

[3]Hamburg Center for Health Economics, University of Hamburg, Hamburg, Germany

[4]Laboratorio Management e Sanità, Institute of Management and Department EMbeDS, Scuola Superiore Sant'Anna, Pisa, Italy

[5]Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada

[*]these authors contributed equally

Corresponding Author:
Erica Barbazza, MSc
Department of Public and Occupational Health, Amsterdam UMC
Amsterdam Public Health Research Institute
University of Amsterdam
Meibergdreef 9
Amsterdam, 1105AZ
Netherlands
Phone: 45 31121711
Email: e.s.barbazza@amsterdamumc.nl

Abstract

Background: Public web-based COVID-19 dashboards are in use worldwide to communicate pandemic-related information. Actionability of dashboards, as a predictor of their potential use for data-driven decision-making, was assessed in a global study during the early stages of the pandemic. It revealed a widespread lack of features needed to support actionability. In view of the inherently dynamic nature of dashboards and their unprecedented speed of creation, the evolution of dashboards and changes to their actionability merit exploration.

Objective: We aimed to explore how COVID-19 dashboards evolved in the Canadian context during 2020 and whether the presence of actionability features changed over time.

Methods: We conducted a descriptive assessment of a pan-Canadian sample of COVID-19 dashboards (N=26), followed by an appraisal of changes to their actionability by a panel of expert scorers (N=8). Scorers assessed the dashboards at two points in time, July and November 2020, using an assessment tool informed by communication theory and health care performance intelligence. Applying the nominal group technique, scorers were grouped in panels of three, and evaluated the presence of the seven defined features of highly actionable dashboards at each time point.

Results: Improvements had been made to the dashboards over time. These predominantly involved data provision (specificity of geographic breakdowns, range of indicators reported, and explanations of data sources or calculations) and advancements enabled by the technologies employed (customization of time trends and interactive or visual chart elements). Further improvements in actionability were noted especially in features involving local-level data provision, time-trend reporting, and indicator management. No improvements were found in communicative elements (clarity of purpose and audience), while the use of storytelling techniques to narrate trends remained largely absent from the dashboards.

Conclusions: Improvements to COVID-19 dashboards in the Canadian context during 2020 were seen mostly in data availability and dashboard technology. Further improving the actionability of dashboards for public reporting will require attention to both technical and organizational aspects of dashboard development. Such efforts would include better skill-mixing across disciplines, continued investment in data standards, and clearer mandates for their developers to ensure accountability and the development of purpose-driven dashboards.

(J Med Internet Res 2021;23(8):e30200) doi:10.2196/30200

KEYWORDS

COVID-19; performance measures; health information management; dashboards; public reporting of health care data; qualitative research; public health; medical informatics; surveillance; communication; assessment; Canada; decision-making; dynamic; development

Introduction

The public reporting of data during a pandemic is a core government function to protect population health and safety [1-3]. It is also critical for fostering accountability, ensuring transparency, and supporting individuals in making informed decisions [4-6]. Unlike past pandemics, COVID-19 has been monitored globally in real-time, resulting in unprecedented collection, analysis, and dissemination efforts.

Public web-based COVID-19 dashboards, as a dynamic means to visually display information at a glance [7], have surged as a popular approach for sharing pandemic-related information. Dashboards are powerful vehicles for communication; the Johns Hopkins Coronavirus Resource Center dashboard [8] reported more than 1 billion interactions per day by April 2020 [9]. However, without careful indicator selection and data collection, analysis, and visualization, dashboards have the potential to mislead, misinform, and incite panic [10,11], or simply to be ignored [12].

In the first half of 2020, our international research network of European and Canadian professionals in health care performance intelligence [13] launched a global study of COVID-19 dashboards. It assessed 158 dashboards from 53 countries in July 2020. It also explored what makes dashboards *actionable*, whereby actionability refers to a dashboard's potential to inform decision-making by the intended users [14]. More specifically, to be actionable, the information should be both *fit for purpose* (meeting a specific information need) and *fit for use* (placing the right information into the right hands at the right time and in a manner that can be understood) [14]. Only 12.7% (20/158) of dashboards evaluated in the mid-2020 study were found to be highly actionable. Seven actionability features were identified among them [15].

Due to the speed at which the dashboards were first launched, traditional technical and organizational aspects of development cycles were cut short [16]. While the urgency of reporting took precedent in the early stages, dashboards are designed to be flexible and continuously iterated. Studies also emphasized the importance of frequent reviews to ensure a dashboard's sustained relevance and use [16,17]. As our initial study was merely a snapshot of the early stages of the pandemic, the extent to which COVID-19 dashboards evolved over a longer period was beyond its scope.

Canada provides a relevant context for further investigating the evolution of COVID-19 dashboards for several reasons. First, public health is the remit of federal, provincial or territorial (PT), and local health authorities [18], which, together with PT ministries, are involved in pandemic monitoring and reporting. This was already reflected in Canada's 2018 multiactor pandemic preparedness plans (for influenza) [19]. In addition to those varied public actors, independent initiatives and the media have also leveraged open data sources in order to generate public-facing COVID-19 dashboards. The range in the types of organizations and their different target geographies of reporting have resulted in a diverse Canadian dashboard landscape.

Second, Canada's experience with COVID-19 intensified in the course of 2020, with an initial peak in early May (about 2500 daily cases) and second peak in November (about 8000 daily cases) [20]. Cases spread to areas of Canada previously untouched by the virus [21]. As a result, the demand for dashboards that provide effective communication and support data-driven decision-making increased throughout the year.

Third, Canadian dashboards were criticized early on for possible information blind spots, including a failure to report race-based data and other social determinants [22,23], as well as for presenting highly aggregated data at the PT level [10,24,25]. The extent to which such limitations persisted into the second half of 2020 is yet to be assessed.

This study explores (1) how public web-based COVID-19 dashboards in the Canadian context evolved in 2020 and (2) whether dashboard actionability increased over time.

Methods

Study Design

Our study adheres to the Standards for Reporting Qualitative Research [26]. We applied qualitative methods comprising (1) a descriptive assessment applying an existing tool [15] for the purposes of systematically and comparatively depicting COVID-19 dashboards; and (2) an expert appraisal using the nominal group technique [27,28] to score the actionability of the dashboards. The study draws on the global sample of 158 dashboards examined in the study by Ivanković et al [15], now confining the focus to dashboards reporting on COVID-19 in the Canadian context (N=26). Importantly, we extended data collection for this sample by collecting data at a second time point, in order to analyze changes between July 2020 (initial assessment) and November 2020 (second assessment). Subsequently, we evaluated the presence of the actionability features identified in the study by Ivanković et al [15] across the sample for both time points.

Panel of Scorers

Data collection was conducted by a panel of eight scorers (EB, DI, SW, KJG, MP, CW, NL, and VB). The panel (four women and four men) aligned with the scorers assembled by Ivanković et al [15] so as to ensure consistency between assessments. The scorers were drawn from an existing international research network of Canadian, European, Latin American, and Asian researchers, each conducting their doctoral research on health care performance intelligence [13]. All scorers had common expertise and training in dealing with health care performance data and in the use of such data for management and governance, as well as prior training and experience with the study's assessment tool. The panel's composition also included French-language competencies (CW) and prior professional policy and research experience in the Canadian context (EB, DI, SW, KJG, MP, and VB).

Assessment Instruments

An assessment tool developed, piloted, and validated by Ivanković et al [15] was applied. The tool assesses COVID-19 dashboards in terms of their purpose and users ("why"), content and data ("what"), and analyses and displays ("how"). Table 1 summarizes the considerations assessed. These derive from communication sciences (the 1948 Lasswell model [29]), the health care performance intelligence discipline [14], earlier studies on the public reporting of health performance data and provision of dashboards in the health domain [30-34], and guidance for reporting during public health crises from the World Health Organization (WHO) [1]. The tool also aligns with existing instruments to measure the quality of health information on the internet [35,36].

We operationalized the appraisal of a dashboard's actionability by drawing on the seven features of highly actionable COVID-19 dashboards, as identified in the study by Ivanković et al [15] (see Table 1). A scoring tool was developed (see Multimedia Appendix 1) to evaluate each feature on a 3-point ordinal scale, scored as "present," "somewhat present," or "not present."

Table 1. Overview of considerations by the method applied.

Method	Instrument	Considerations assessed/scored: guiding questions/statements
Descriptive assessment	Assessment tool[a]	• Purpose and audience: Is the purpose and audience mentioned? • Indicator themes: What indicators are reported on? • Data: Are data sources and metadata specified? • Types of analysis: Does the analysis include time trends, and geographic and population break downs? • Presentation: How is data visualized, interpreted, simplified, and interacted with?
Expert appraisal	Seven features of highly actionable dashboards-scoring tool[b]	• Know the audience and their information needs: The intended audience and their information needs are known and responded to. • Manage the type, volume, and flow of information: The type, volume, and flow of information on the dashboard are well managed. • Report data sources and methods clearly: The data sources and methods for calculating values are made clear. • Link time trends to policy decisions: Information is reported over time and contextualized with policy decisions made. • Provide data "close to home": Data are reported at relevant geographic break downs. • Break down the population to relevant subgroups: Data are reported by relevant population subgroups. • Use storytelling and visual cues: Brief narratives and visual cues are used to explain the meaning of data.

[a] Refer to the study by Ivanković et al [15] for the full assessment tool.

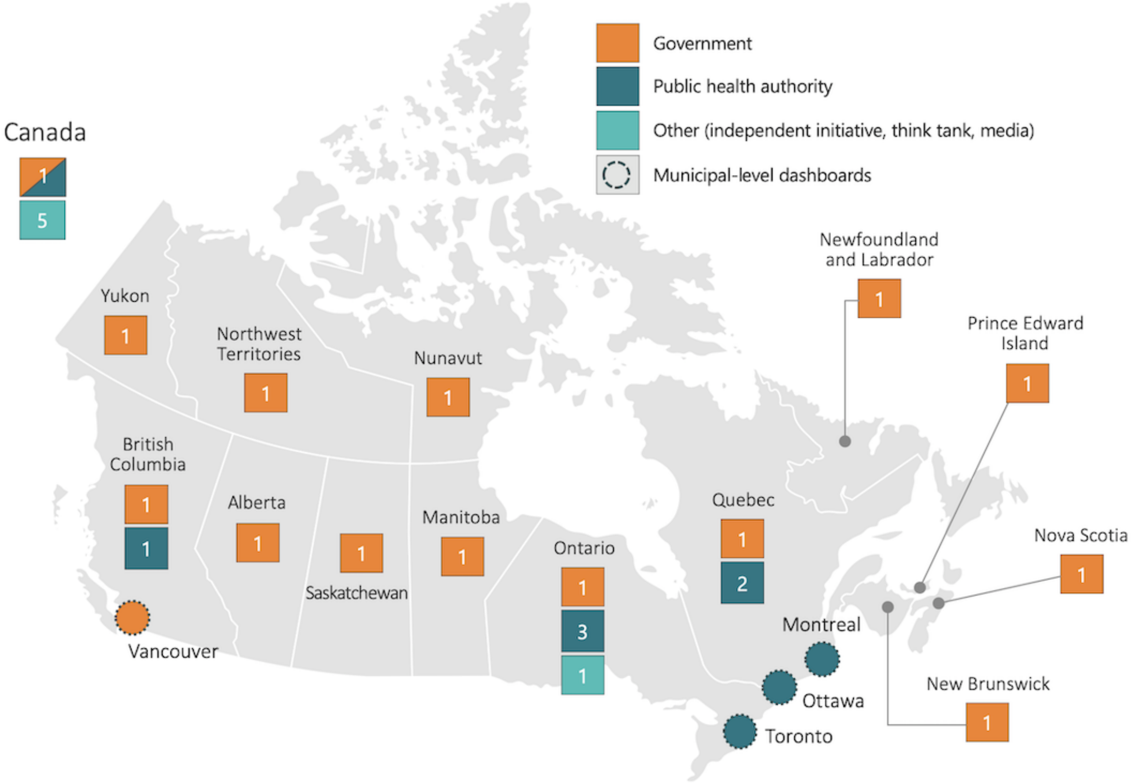
[b] Refer to Multimedia Appendix 1 for the full scoring tool.

Study Sample

COVID-19 dashboards for sample inclusion were determined on the basis of the following three criteria: (1) the reporting of key performance indicators related to COVID-19; (2) the use of some form of visualization; and (3) availability in an online web-based format. It means password-protected COVID-19 dashboards for internal use by public authorities were excluded from this study. No restrictions were imposed in terms of a dashboard's primary level of reporting (eg, national, regional, and local) or the type of organization responsible for its development (eg, government, academia, news or media, industry, and private initiative). Sampling was conducted from May 19 to June 30, 2020, and involved searches of COVID-19 policy monitoring platforms (eg, the North American COVID-19 Policy Response Monitor [37]) and of research reports (eg, a June 2020 pan-Canadian catalogue of governmental COVID-19 dashboards [38]), as well as expert recommendations from researchers actively engaged in the COVID-19 response, who were contacted via email. In total, 31 dashboards reporting on the Canadian context were identified, five of which were duplicates and excluded from further analysis. Further details about the sampling are mentioned in the study by Ivanković et al [15].

The final sample (N=26) included dashboards reporting at the national level (n=6), PT level (n=16) (including at least one from each of Canada's 13 provinces and territories), and municipal level (n=4), capturing reporting from the capital (Ottawa) and the three largest cities (Montreal, Toronto, and Vancouver). Figure 1 maps the pan-Canadian distribution and the variations in the types of organizations responsible for developing the dashboards. These included federal or PT governments (14/26, 54%), public health authorities (6/26, 23%), and others (6/26, 23%), including independent initiatives (eg, #HowsMyFlattening and COVID-19 Canada Open Data Working Group), industry (eg, Esri and Deloitte), and media (Canadian Broadcasting Corporation). See Multimedia Appendix 2 for the complete list of dashboards.

Figure 1. Distribution of COVID-19 dashboards sampled and types of organizations responsible for their development. Circles denote municipal-level dashboards included in the sample, and the colors denote the respective organization types. These dashboards are counted in the tally shown per jurisdiction. The Public Health Agency of Canada's COVID-19 dashboard is hosted on the federal Government of Canada webpage. In other instances, dashboards developed by public health authorities are hosted on dedicated webpages.


Government
Public health authority
Other (independent initiative, think tank, media)
Municipal-level dashboards
Canada
1
5
Yukon
1
Northwest Territories
1
Nunavut
1
Newfoundland and Labrador
1
Prince Edward Island
1
British Columbia
1
1
Alberta
1
Saskatchewan
1
Manitoba
1
Ontario
1
3
1
Quebec
1
2
Nova Scotia
1
Vancouver
Montreal
Ottawa
Toronto
New Brunswick
1

Descriptive Assessment

Each dashboard was assessed in English or French. The assessments were limited to a dashboard's main page and to content accessible within one interaction (click). This approach was designed to increase consistency in the content evaluated, and it enabled us to gauge the dashboard's prioritization and hierarchy of content. Archives were generated to create a record of each dashboard on the date reviewed (see Multimedia Appendix 2). Dashboards were distributed among the scorers as described in the study by Ivanković et al [15]. This distribution (averaging three dashboards per scorer) remained consistent between time points as follows: the same scorers assessed the same dashboards in both July and November 2020. All assessments additionally underwent reviews by the first authors (EB and DI) to verify completeness and consistency.

Expert Appraisal

To assess the presence of the seven defined features of highly actionable COVID-19 dashboards, we organized a series of three-person panels, involving the original scorer of each dashboard joined by two other experts (the first authors or another panel member), in December 2020. Prior to the start of the appraisal by each panel, a workshop with the scorers was organized to calibrate the approach to scoring.

Scoring was informed by the original data records and archives generated in the two descriptive assessments (July and November 2020). Importantly, each of the seven actionability features were appraised with consideration to the dashboard's stated or inferred purpose and audience. It means the appraisal of each feature differentiated between the intended use of the dashboard by national, PT, or municipal general public audiences, unless further specified. In line with the nominal group technique approach [27,28], the three panel members first independently scored the presence of each feature on the dashboard using the scoring tool described above. The proportion of identical ratings for each dashboard was calculated, and virtual panel discussions were convened between the three scorers involved [39,40].

Prior to those discussions, partial or full agreement (two- or three-way consensus) had been reached on 83.5% (304/364) of the items scored, with full three-way agreement on 50.0% (182/364) (see Multimedia Appendix 3). During the panel discussions, all items without full agreement were debated. All panels reached final agreement by discussion or re-examining the data records or archives.

Data Analysis

We used descriptive statistics to analyze the data at the two time points. We first determined the number and percentage of dashboards in which each item (ie, each consideration) of the descriptive assessment had been recorded as present in the July or November assessment or both. The net change for each item was calculated as the change in the total number of dashboards and the direction of that change between time points. To analyze score changes for the actionability features, we calculated feature-by-feature totals in both July and November, applying a 3-point ordinal scale (not present, somewhat present, and present). Using the same approach applied to analyze changes over time in the descriptive assessments, we calculated the net change per feature as the change in the total number of positively scored dashboards, noting the direction of that change.

For free-text fields in the descriptive assessment tool, we used both deductive and inductive thematic analysis to identify themes [41,42]. This applied to responses on considerations such as a dashboard's purpose of use and audience, indicator titles, and considerations with "other" as an answer category. Topics explored in the assessment tool were used to guide the deductive thematic analysis. In analyzing the titles of indicators reported by the dashboards, we applied the existing WHO classification of types of pandemic-related information. Indicators were analyzed by the types of information as follows: public health and epidemiology, health system management, social and economic impact, and behavioral insights [1]. Given the observed variability in the phrasing of indicator titles, the first authors grouped key performance indicators by themes. New themes that emerged were identified using an inductive approach.

Ethics Approval

This study involved the analysis of publicly available COVID-19 dashboards. Ethics approval was not required.

Results

Sampled Dashboards

The 26 Canadian COVID-19 dashboards were assessed in the time frames July 7 to July 20 and November 23 to December 2, 2020, with an average of 135 days between assessments (range 132-140). All dashboards remained active, with regular, typically daily updating, aside from one (City of Vancouver), which was still accessible but last updated in August 2020. As expected, given the wide differences in population size and density across Canadian provinces and territories, the cumulative number of COVID-19 cases reported by the dashboards for their respective geographic areas ranged from 0 cases in Nunavut to more than 55,000 in Quebec in July, and from 15 cases in Northwest Territories to more than 140,000 in Quebec in November. Cumulative numbers of COVID-19 cases and deaths on the assessment dates are reported in Multimedia Appendix 2.

Changes to the Dashboards Over Time

Table 2 reports how the dashboards changed over time according to the descriptive assessment. The changes can be summarized as presented below.

Table 2. Description of changes to Canadian COVID-19 dashboards (N=26) over time in 2020.

Consideration and description	July value, n (%)	November value, n (%)	Net change[a]
Purpose and audience			
Purpose: Purpose of use of the dashboard stated	10 (39%)	10 (39%)	0
Audience: Intended audience (user) stated	3 (12%)	4 (15%)	+1
Indicator themes			
Spread and death			
Cases (all confirmed cases)	25 (96%)	25 (96%)	0
Deaths	20 (77%)	21 (81%)	+1
Recovered (healed, cured)	17 (65%)	18 (69%)	+1
Active cases	12 (46%)	12 (46%)	0
Mortality rate (case fatality rate)	4 (15%)	4 (15%)	0
Reproduction rate (attack rate)	1 (4%)	5 (19%)	+4
Testing			
Testing (total number tested, PCR[b] tests)	17 (65%)	19 (73%)	+2
Testing rates (positivity, negative tests)	10 (39%)	15 (58%)	+5
Tests pending results	4 (15%)	2 (8%)	−2
Testing turnaround	0 (0%)	3 (12%)	+3
Risk management			
Self-quarantine (isolation notices)	1 (4%)	1 (4%)	0
Contact tracing	2 (8%)	2 (8%)	0
Hospital care			
Hospitalized (admissions, discharges)	16 (62%)	15 (58%)	−1
Admitted to the ICU[c] (critical condition)	10 (39%)	12 (46%)	+2
On a ventilator	3 (12%)	3 (12%)	0
Health system capacity			
Hospital bed capacity (availability)	2 (8%)	2 (8%)	0
ICU bed capacity	3 (12%)	2 (8%)	−1
Ventilator capacity (available ventilators)	3 (12%)	2 (8%)	−1
Non-COVID-19 service usage	1 (4%)	1 (4%)	0
Personal protective equipment stock	1 (4%)	1 (4%)	0
Economic/social impact			
Employment and hardship relief	4 (15%)	4 (15%)	0
Transport, trade, and international travel	2 (8%)	3 (12%)	+1
Behavioral: Public risk perception/restriction adherence	5 (19%)	3 (12%)	−2
Other			
Future projections (modelling)	1 (4%)	1 (4%)	0
Risk-level/current phase (composite score)	2 (8%)	4 (15%)	+2
Data sources and metadata			
Sources: Data sources are noted	18 (69%)	18 (69%)	0
Metadata: Metadata are specified	11 (42%)	14 (54%)	+3
Types of analyses			
Time trend			

Consideration and description	July value, n (%)	November value, n (%)	Net change[a]
Time trend analysis available	21 (81%)	23 (89%)	+2
Customizable time trend	4 (15%)	10 (39%)	+6
Number of geographic levels			
1 level	6 (23%)	3 (12%)	−3
2 levels	14 (54%)	15 (58%)	+1
3 or more levels	6 (23%)	8 (31%)	+2
Types of geographic levels of analysis			
International	3 (12%)	3 (12%)	0
National	9 (35%)	8 (31%)	−1
Regional (province/territory)	22 (85%)	22 (85%)	0
Health regions	10 (39%)	15 (58%)	+5
Municipal (city)	8 (31%)	8 (31%)	0
Neighborhood (postcode)	3 (12%)	2 (8%)	−1
Disaggregation options			
Age	18 (69%)	17 (65%)	−1
Sex	14 (54%)	15 (58%)	+1
Mode of transmission	5 (19%)	6 (23%)	+1
Long-term care facilities	5 (19%)	5 (19%)	0
Schools	2 (8%)	5 (19%)	+3
Ethnicity	0 (0%)	2 (8%)	+2
Race	0 (0%)	2 (8%)	+2
Comorbidities	1 (4%)	1 (4%)	0
Socioeconomic status	1 (4%)	1 (4%)	0
Health workers	3 (12%)	1 (4%)	−2
Presentation			
Type of visualization			
Table	20 (77%)	25 (96%)	+5
Graph/chart	21 (81%)	22 (85%)	+1
Map	15 (58%)	18 (69%)	+3
Narratives to interpret data			
Yes, to clarify the quality of the data	13 (50%)	18 (69%)	+5
Yes, to clarify the meaning of the data	12 (46%)	11 (42%)	−1
Simplification techniques			
Use of color coding	15 (58%)	15 (58%)	0
Size variation	3 (12%)	4 (15%)	+4
Icons	3 (12%)	7 (27%)	−2
Interactive options			
More information	18 (69%)	18 (69%)	0
Change of information	7 (27%)	10 (39%)	+3

Consideration and description	July value, n (%)	November value, n (%)	Net change[a]
Change of display	5 (19%)	6 (23%)	+1

[a] Net change refers to the total number of dashboards and the direction of overall change between time points. Importantly, no net change (0) can mean both no change or the same number of dashboards increased and decreased for the specific consideration.

[b] PCR: polymerase chain reaction.

[c] ICU: intensive care unit.

Purpose and Audience

There was no change in the extent to which dashboards stated their purpose of reporting, with just over one-third (10/26, 38%) doing so in both July and November. Where stated, the most frequent specific aims of dashboards were to provide simplified information in an “easy-to-digest, actionable way” [43] and to “help prevention strategies reach those people most affected” [44]. The explicit mention of a target audience was even less frequent, being found on just four dashboards (4/26, 15%) in November, a marginal increase from July (3/26, 12%). Target audiences were denoted as “general public,” “businesses,” or “public health leaders.” Notable improvements over time were made by Ontario’s #HowsMyFlattening [43], with the introduction of two dashboard viewing modes (“personal” and “geek”) to serve the information needs of different audiences.

Indicator Themes

Across the dashboards, public health and epidemiological indicators, followed by health system management indicators, were the most frequently reported indicators at both time points. Behavioral and socioeconomic indicators were rare. An average of seven indicator themes were reported per dashboard in November (range 2-17), compared with six in July (range 2-15). Several indicators became more prevalent in November, including viral reproduction rates, testing rates, testing turnaround times, and composite scores. Six dashboards (6/26, 23%) reduced the number of indicator themes reported, most often removing indicators on active cases. In some instances, indicators had been moved from the dashboard to new tabs or pages, as in Ottawa [45], which relocated indicators on behavioral insights to new tabs no longer within direct access of the main dashboard page assessed. Indicators on serology tests, doubling rates, and testing stock, which had been present on dashboards previously assessed internationally [15], were not reported at either time point on the sampled dashboards.

Data Sources and Metadata

A third (8/26, 31%) of the dashboards, all government developed, did not explicitly report data sources in July or November. Dashboards typically drew data from jurisdiction-specific health services and public health authorities, hospital databases, and, for comparisons with other countries, the Johns Hopkins University Coronavirus Resource Center dashboard. Dashboards reporting metadata (supplementary details on the calculation of the indicators) increased to more than 50% (14/26, 54%) by November (from 11/26, 42%, in July). Notably, the COVID-19 in Canada dashboard published a detailed technical report on its data set produced by the COVID-19 Canada Open Data Working Group initiative [46,47].

Types of Analyses

A slight increase in the number of dashboards reporting time-trend data was observed between July and November (from 21/26, 81% to 23/26, 88%). Improvements were also made to the availability of customizable time scales, allowing users to zoom in on specific time frames of interest (from 4/26, 15% to 10/26, 38%).

Modifications were made to report subregional geographic breakdowns of data, with more than half (15/26, 58%) of the dashboards including breakdowns by health regions in November, as compared with 10 (10/26, 38%) in July. Age and sex remained the most common population breakdowns in November (17/26, 65%, as against 15/26, 58%, in July), followed by mode of transmission (6/26, 23%) and long-term care facilities (5/26, 19%). Schools emerged as a new type of breakdown in November, though present on only one-fifth (5/26, 19%) of dashboards.

Presentation

Between July and November, most dashboards slightly improved the number and variety of chart types, simplification techniques, and interactive features they made available. This was mostly done by introducing maps or additional tables and icons, as well as user-directed modifications to the information displayed. New features that emerged in November included options to subscribe to email updates for alerts (eg, #HowsMyFlattening [43] and Ottawa [45]). Two dashboards (Quebec [48] and Ontario [49]) introduced user feedback surveys.

Text providing details on data quality was present on more than two-thirds (18/26, 69%) of dashboards in November, compared with half (13/26, 50%) in July. For example, Esri’s dashboard included lay-language explanations of values with statements such as “*Why do I sometimes see negative numbers?* Some values reported (like total cases) are cumulative. They always go up. Other values (like hospitalizations) fluctuate and can go up or down day-to-day” [50]. Narratives to explain the meaning of statistics and trends were provided by fewer than half (11/26, 42%) of the dashboards in November. Explanations of trends and their meaning included the following description provided by the COVID-19 in Canada dashboard: “Graphs display trends for daily cases and deaths over time on a logarithmic scale. An upward slope means the number of cases/deaths reported each day is still growing. A flat line means the number of cases/deaths reported each day is staying the same. A downward slope means the number of cases/deaths reported each day is falling” [20].

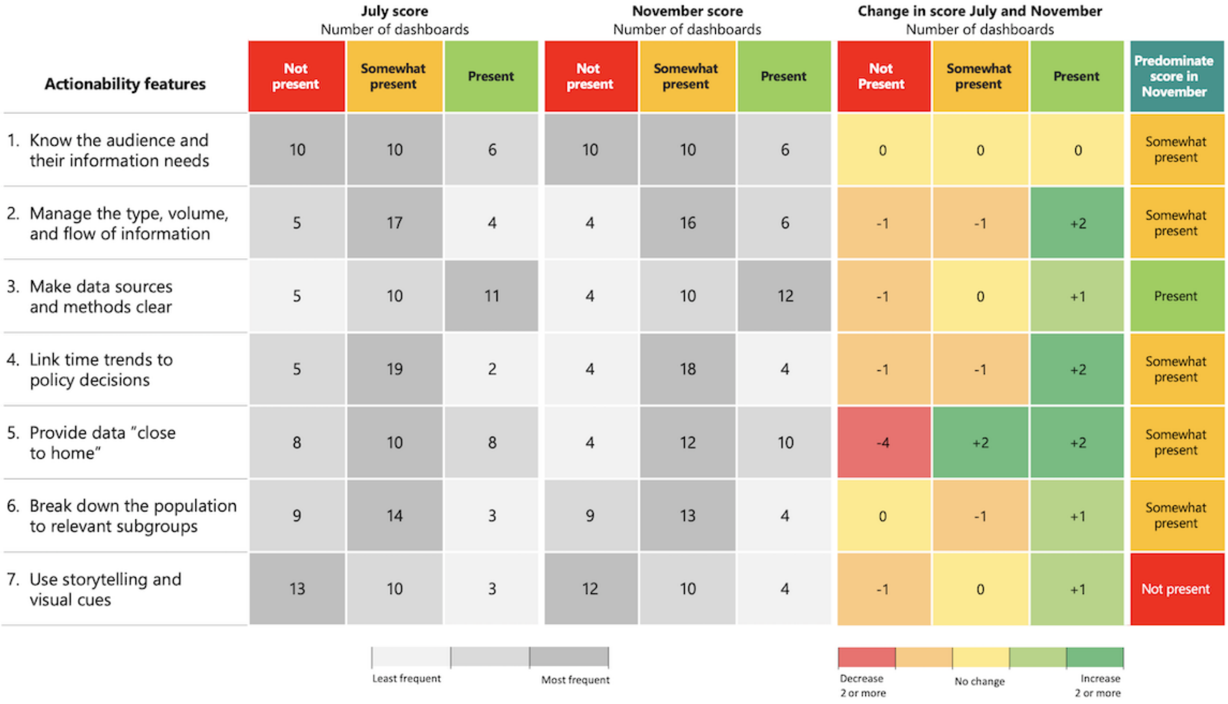
Actionability Features Over Time

Of the 26 dashboards assessed, none was found to fully present all seven of the defined actionability features either in July or November. Overall, 8% (2/26) of dashboards were assessed in

July as having five or more actionability features fully present, doubling to 15% (4/26) of dashboards in November. Three quarters (77%, 20/26) of dashboards had two or fewer features fully present in July and 65% (17/26) had two or fewer features fully present in November. Seven dashboards increased their score of fully present features. Although two dashboards scored lower in November, the decrease was largely attributable to modifications in the type of information reported on the main dashboard page, as indicators were moved to other dedicated pages.

The actionability feature most widely present on dashboards in both July and November was the clarity of data sources and methods, while the use of storytelling and visual cues was the feature most frequently absent (Figure 2). Among the seven defined features of actionability, improvements were observed in all but one (knowing the audience and their information needs), which was present on fewer than a quarter of the dashboards at either time point. Improvements were most pronounced for the feature involving geographic breakdown, with average scores increasing by nearly a quarter from July to November. Second to these improvements were improvements in the use of time trends, although explicit links between data and policy decisions and infection control measures remained infrequent.

Figure 2. Change in actionability across dashboards (n=26) over time in 2020. Not present: the feature is not found on the dashboard; somewhat present: some elements of the feature are present on the dashboard but room for improvement; present: the specific feature is clearly demonstrated and a good practice example of the feature is present. See Multimedia Appendix 1 for full scoring details and Multimedia Appendix 3 for the level of agreement between panel members.



Actionability features	July score Number of dashboards			November score Number of dashboards			Change in score July and November Number of dashboards			
	Not present	Somewhat present	Present	Not present	Somewhat present	Present	Not Present	Somewhat present	Present	Predominate score in November
1. Know the audience and their information needs	10	10	6	10	10	6	0	0	0	Somewhat present
2. Manage the type, volume, and flow of information	5	17	4	4	16	6	-1	-1	+2	Somewhat present
3. Make data sources and methods clear	5	10	11	4	10	12	-1	0	+1	Present
4. Link time trends to policy decisions	5	19	2	4	18	4	-1	-1	+2	Somewhat present
5. Provide data "close to home"	8	10	8	4	12	10	-4	+2	+2	Somewhat present
6. Break down the population to relevant subgroups	9	14	3	9	13	4	0	-1	+1	Somewhat present
7. Use storytelling and visual cues	13	10	3	12	10	4	-1	0	+1	Not present

Discussion

Principal Findings

In this study, we explored changes made in the course of 2020 to public web-based COVID-19 dashboards in Canada and appraised their actionability for decision-making purposes. Although the dashboards we sampled varied in their specific geographic focuses, they all shared an increasing relevance in supporting data-driven decision-making in their respective audiences as the severity of the COVID-19 pandemic intensified across the country. Broadly speaking, from the perspective of the health care performance intelligence we applied, we observed that subtle improvements were made to the dashboards between July and November 2020. Improvements were most pronounced with regard to dashboard technology solutions (better customizable time trends, and new charts and graphs) and data provision (new indicators, more transparency on metadata, and more geographic granularity). Modifications to further develop communicative elements were less pronounced or even absent during the period assessed. These results were mirrored in the scoring of actionability features.

COVID-19 dashboards worldwide are powered by a somewhat common range of software service providers (eg, ArcGIS, Tableau, and Power BI). We presume that some improvements observed across our sample can be credited to new technical features rolled out by such providers during 2020. For example, the use of adjustable time trends was a feature introduced on more than a third of the dashboards by November and was evidently an added element in the underlying software. However, while the industry may be credited with spearheading the technical development of dashboards, the current practice from a technological perspective of measuring actionability through *user clicks* [51] exposes some limitations. To give an example, the enhanced sophistication of the technology behind more interactive time trends used on dashboards was not complemented with improvements to incorporate the enactment of policy restrictions into time-trend graphs so as to visualize subsequent effects of those restrictions. This was despite the merits of such a visualization [15] and the fact that such a technique was already being applied on dashboards in countries like Australia [52] and Slovenia [53]. In our sample, we did observe dashboards that excelled in actionability, successfully leveraging the skills of specialists in technology, data, public health, and communication [43,54]. This finding is consistent with the findings in previous studies that have shown the importance of diverse stakeholder engagement for achieving actionable performance measurement, data reporting, and dashboard use [55-57]. In future research, we intend to further explore the perspective of dashboard developers, including their team profiles.

Improved geographic granularity and transparency of methods may be supported by initiatives like the COVID-19 Canada Open Data Working Group [20]. The overall subtlety of changes in available data and its specificity might be a symptom of underlying system barriers, in particular in relation to the collection and reporting of disaggregated data [58]. Researchers in the Canadian context have called attention to data management issues arising from unharmonized privacy laws, public/private data custodianship, and obstacles to the reuse of data for research [59]. The collection of race-based data in Canada is fragmented [60], and a pan-Canadian standard was proposed only in July 2020 [61]. There is a responsibility to act in cases where missing data could be masking inequitable burdens of the pandemic [62,63]. The potential equity-promoting impact of subpopulation-based approaches to the analysis and use of data has already been highlighted in Toronto [64]. Countries that report race- and ethnicity-based COVID-19 data, like New Zealand [65] and the United States [66], may be a source of insights into necessary data governance standards, privacy protections, and data infrastructure.

Our findings also reveal a responsiveness to the evolving nature of the pandemic, with multiple dashboards adding school cases or outbreaks as a data disaggregation option and turnaround times for virus testing as an indicator. Shortly after our second assessment, many dashboards also began reporting on vaccinations. Less advanced dashboards, from areas not seriously affected by the pandemic in the spring of 2020, made considerable progress in the second half of the year, as COVID-19 became more widespread. While such changes confirm that dashboards continued developing with time, the clarity of their intended aims and audiences nevertheless remained an underdeveloped attribute, despite wide recognition of the fundamental importance of data driven by a clear purpose and information need [14,67-70]. This may be a symptom of data governance constraints or, more specifically, of unclear responsibilities and mandates delegated to developers, as evidenced by the multiple public actors (eg, PT governments and PT public health authorities) that were reporting on the same geographies with nearly equivalent content. Although COVID-19 dashboards began as a need-based short-term tool for monitoring and communicating on the pandemic, this function has evolved with time. Dashboards must now face the mid-term challenge of dual-track health system monitoring, reporting both on the pandemic and on non-COVID health care [71], as well as the long-term challenge of integration into standard health system performance measurement. Rethinking the development of dashboards governed by clear mandates will be essential to ensure that relevant high-quality information is transparently delivered to well-defined audiences.

Strengths and Limitations

To our knowledge, this is the first study to comparatively explore and critically reflect on changes to COVID-19 dashboards over time from a health care performance intelligence perspective. The study was enriched by the expertise of the panel, whose members had prior experience in assessing COVID-19 dashboards internationally, as well as a shared reflexive lens to gauge both the technical and communication aspects of the dashboards. Additionally, given the sustained relevance of COVID-19 dashboards, our findings are pertinent to both short-term improvements in COVID-19 dashboards and their longer-term utility in addressing future public health crises.

We acknowledge several limitations. First, the stages of the pandemic and its severity varied considerably across our sample, possibly contributing to differences with respect to the data available and the prioritization of a dashboard’s development. Despite this, the general direction of change was found to be common, averaging a three-fold increase in COVID-19 cases across locations between our assessment time points (see Multimedia Appendix 2). Second, the expert-based appraisal of actionability we employed is not a guaranteed reflection of a dashboard’s use in practice. The first-hand experiences of dashboard users merit further study to obtain practical real-world insights that can complement the concepts explored here. Third, our archiving of dashboards was limited to their main page. Dashboards with multiple tabs could therefore not be revisited in full for scoring purposes. To minimize the potential loss of information, all dashboards were assessed and evaluated by the same scorer in both July and November. Lastly, to permit comparisons over time, our sample was limited to dashboards identified in our search in May 2020. Any new dashboards that followed would have been missed. An exhaustive sample was beyond the study’s aims; however, we achieved geographic representativeness, as well as reasonable diversity in level (national, jurisdictional, and municipal) and in the types of providing organizations.

Conclusion

Actionable dashboards are needed to enable effective decision-making across audiences. Dashboards are tools of continuing importance during the COVID-19 pandemic, but sustaining their actionability requires responsiveness to the pandemic’s stages. Improvements made to COVID-19 dashboards in the Canadian context from July to November 2020 appear to be driven mainly by certain technological and data improvements. The effective use of communication features remained underdeveloped at both points in time. COVID-19 dashboard developers need to better leverage the expertise of public health and communication specialists, in order to ensure that data will truly become information that is readily accessible and relevant to a public audience. Strategic system improvements to prioritize data standards, for example, those with respect to subpopulation-based data, are needed to achieve more significant gains in actionability. As the pandemic continues to evolve, attention will need to shift toward converting dashboards from their initial status as temporary monitoring and communication tools into instruments that are integrated into routine health system performance monitoring. Accomplishing that will also require improved governance arrangements that clarify roles and responsibilities. In the short term, continued improvements are urgently needed with respect to all seven of the identified actionability features, in order to make COVID-19 dashboards more fit for their purpose and use.

Acknowledgments

The authors thank and recognize the full team involved in the work on the international sample of COVID-19 dashboards that preceded this study. We also thank the reviewers who commented on an earlier version of the paper and Michael Dallas for language editing. This work was carried out by the Marie Skłodowska-Curie Innovative Training Network Healthcare Performance Intelligence Professionals (HealthPros) that has received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement number 765141.

Authors' Contributions

EB and DI contributed equally as co-first authors. EB, DI, SA, NK, and DK designed the study. Data collection was conducted by EB, DI, SW, KJG, MP, CW, NL, and VB. EB and DI drafted the manuscript. All authors revised the article, gave final approval for the version to be published, and agreed to be accountable for all aspects of the work.

Conflicts of Interest

None declare.

Multimedia Appendix 1
Scoring tool on actionability features.
[DOCX File , 20 KB - jmir_v23i8e30200_app1.docx]

Multimedia Appendix 2
Overview of Canadian COVID-19 dashboards assessed.
[DOCX File , 27 KB - jmir_v23i8e30200_app2.docx]

Multimedia Appendix 3
Scoring distribution and extent of agreement prior to joint workshops.
[DOCX File , 15 KB - jmir_v23i8e30200_app3.docx]

References

1. Strengthening and adjustingpublic health measures throughout the COVID-19 transition phases: policyconsiderations for the WHO European Region, 24 April 2020. World Health Organization. 2020. URL: https://apps.who.int/iris/handle/10665/332467 [accessed 2021-07-31]
2. Frieden TR. Government's role in protecting health and safety. N Engl J Med 2013 May 16;368(20):1857-1859. [doi: 10.1056/NEJMp1303819] [Medline: 23593978]
3. UN/DESA Policy Brief #61: COVID-19: Embracing digital government during the pandemic and beyond. United Nations. 2020. URL: https://www.un.org/development/desa/dpad/publication/un-desa-policy-brief-61-covid-19-embracing-digital-government-during-the-pandemic-and-beyond/ [accessed 2021-07-31]
4. Strengthening the health systems response to COVID-19: policy brief: recommendations for the WHO European Region (1 April 2020) . World Health Organization. 2020. URL: https://apps.who.int/iris/handle/10665/333072 [accessed 2021-07-31]
5. Infodemic Management. World Health Organization. URL: https://www.who.int/teams/risk-communication/infodemic-management [accessed 2021-07-31]
6. Flattening the COVID-19 peak: containment and mitigation policies. OECD. 2020. URL: https://read.oecd-ilibrary.org/view/?ref=124_124999-yt5ggxirhc&title=Flattening_the_COVID-19_peak-Containment_and_mitigation_policies [accessed 2021-07-31]
7. Ghazisaeidi M, Safdari R, Torabi M, Mirzaee M, Farzi J, Goodini A. Development of Performance Dashboards in Healthcare Sector: Key Practical Issues. Acta Inform Med 2015 Oct;23(5):317-321 [FREE Full text] [doi: 10.5455/aim.2015.23.317-321] [Medline: 26635442]
8. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. ArcGIS Online. URL: https://www.arcgis.com/apps/dashboards/bda7594740fd40299423467b48e9ecf6 [accessed 2021-07-31]
9. Perkel J. Behind the Johns Hopkins University coronavirus dashboard. Nature Index. 2020. URL: https://www.natureindex.com/news-blog/behind-the-johns-hopkins-university-coronavirus-dashboard [accessed 2021-07-31]

10. Rocha R. The flurry of daily pandemic data can be overwhelming. Here's how to make sense of it. CBC. 2020. URL: https://www.cbc.ca/news/health/covid-19-pandemic-data-primer-stats-charts-1.5513222 [accessed 2021-07-31]
11. Tomes N. Managing the modern infodemic. CMAJ 2020 Oct 26;192(43):E1311-E1312 [FREE Full text] [doi: 10.1503/cmaj.201905] [Medline: 33106305]
12. Shelton T. A post-truth pandemic? Big Data & Society 2020 Oct 21;7(2):205395172096561. [doi: 10.1177/2053951720965612]
13. International Training Network for Healthcare Performance Intelligence Professionals. HealthPros. 2020. URL: https://www.healthpros-h2020.eu/ [accessed 2021-07-31]
14. Barbazza E, Klazinga NS, Kringos DS. Exploring the actionability of healthcare performance indicators for quality of care: a qualitative analysis of the literature, expert opinion and user experience. BMJ Qual Saf 2021 May 07:bmjqs-2020-011247 [FREE Full text] [doi: 10.1136/bmjqs-2020-011247] [Medline: 33963072]
15. Ivanković D, Barbazza E, Bos V, Brito Fernandes, Jamieson Gilmore K, Jansen T, et al. Features Constituting Actionable COVID-19 Dashboards: Descriptive Assessment and Expert Appraisal of 158 Public Web-Based COVID-19 Dashboards. J Med Internet Res 2021 Feb 24;23(2):e25682 [FREE Full text] [doi: 10.2196/25682] [Medline: 33577467]
16. Weggelaar-Jansen AMJWM, Broekharst DSE, de Bruijne M. Developing a hospital-wide quality and safety dashboard: a qualitative research study. BMJ Qual Saf 2018 Dec;27(12):1000-1007 [FREE Full text] [doi: 10.1136/bmjqs-2018-007784] [Medline: 29950323]
17. Yigitbasioglu OM, Velcu O. A review of dashboards in performance management: Implications for design and research. International Journal of Accounting Information Systems 2012 Mar;13(1):41-59. [doi: 10.1016/j.accinf.2011.08.002]
18. Marchildon GP, Allin S. Chapter 5: Provision of services. In: Merkur S, editor. Canada: Health System Review 2020: Health Systems in Transition. Copenhagen, Denmark: WHO Regional Office for Europe Copenhagen; 2020:100-129.
19. Canadian Pandemic Influenza Preparedness: Planning Guidance for the Health Sector. Public Health Agency of Canada. 2018. URL: https://www.canada.ca/en/public-health/services/flu-influenza/canadian-pandemic-influenza-preparedness-planning-guidance-health-sector.html [accessed 2021-07-31]
20. COVID-19 Canada Open Data Working Group: Monitoring Canada's COVID-19 epidemic. COVID-19 Canada Open Data Working Group. 2021. URL: https://opencovid.ca/ [accessed 2021-07-31]
21. Coronavirus disease (COVID-19): Outbreak update. Government of Canada. 2020. URL: https://www.canada.ca/en/public-health/services/diseases/2019-novel-coronavirus-infection.html#a7 [accessed 2021-07-31]
22. McKenzie K. Race and ethnicity data collection during COVID-19 in Canada: if you are not counted you cannot count on the pandemic response. Royal Society of Canada. 2020. URL: https://rsc-src.ca/en/race-and-ethnicity-data-collection-during-covid-19-in-canada-if-you-are-not-counted-you-cannot-count [accessed 2021-07-31]
23. Bain B, Dryden O, Walcott R. COVID-19 discriminates against Black lives via surveillance, policing and lack of data: U of T experts. University of Toronto. 2020. URL: https://www.utoronto.ca/news/covid-19-discriminates-against-black-lives-surveillance-policing-and-lack-data-u-t-experts [accessed 2021-07-31]
24. Hansen G, Cyr A. Canada's Decentralized "Human-Driven" Approach During the Early COVID-19 Pandemic. JMIR Public Health Surveill 2020 Dec 23;6(4):e20343 [FREE Full text] [doi: 10.2196/20343] [Medline: 33315582]
25. Matukas LM, Dhalla IA, Laupacis A. Aggressively find, test, trace and isolate to beat COVID-19. CMAJ 2020 Oct 05;192(40):E1164-E1165 [FREE Full text] [doi: 10.1503/cmaj.202120] [Medline: 32907821]
26. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014 Sep;89(9):1245-1251 [FREE Full text] [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
27. Jones J, Hunter D. Consensus methods for medical and health services research. BMJ 1995 Aug 05;311(7001):376-380 [FREE Full text] [doi: 10.1136/bmj.311.7001.376] [Medline: 7640549]
28. Fink A, Kosecoff J, Chassin M, Brook RH. Consensus methods: characteristics and guidelines for use. Am J Public Health 1984 Sep;74(9):979-983. [doi: 10.2105/ajph.74.9.979] [Medline: 6380323]
29. Lasswell HD. The structure and function of communication in society. In: Bryson L, editor. The Communication of Ideas. New York, NY: Harper and Row; 1948:136-139.
30. 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]
31. McDowell R. Signs to look for: criteria for developing and selecting fit for purpose indicators. PricewaterhouseCoopers. 2017. URL: https://ourlandandwater.nz/wp-content/uploads/2019/03/IWG_Indicators-thinkpiece-1.pdf [accessed 2021-07-31]
32. Peters E, Dieckmann N, Dixon A, Hibbard JH, Mertz CK. Less is more in presenting quality information to consumers. Med Care Res Rev 2007 Apr;64(2):169-190. [doi: 10.1177/10775587070640020301] [Medline: 17406019]
33. Hibbard JH, Peters E. Supporting informed consumer health care decisions: data presentation approaches that facilitate the use of information in choice. Annu Rev Public Health 2003;24:413-433. [doi: 10.1146/annurev.publhealth.24.100901.141005] [Medline: 12428034]
34. Ballard A. Promoting Performance Information Use Through Data Visualization: Evidence from an Experiment. Public Performance & Management Review 2019 Apr 19;43(1):109-128. [doi: 10.1080/15309576.2019.1592763]

35. Charnock D. The DISCERN Handbook: Quality Criteria for Consumer Health Information on Treatment Choices. Oxford, UK: Radcliffe Medical Press; 1998.
36. SAM: Suitability Assessment of Materials for evaluation of health-related information for adults. Aspirus Library. 2008. URL: http://aspiruslibrary.org/literacy/SAM.pdf [accessed 2021-07-31]
37. North American COVID-19 Policy Response Monitor. University of Toronto. 2020. URL: https://ihpme.utoronto.ca/research/research-centres-initiatives/nao/covid19/ [accessed 2021-07-31]
38. Provincial and Territorial Covid-19 Reporting Dashboards. Public Health Physicians of Canada. URL: http://www.phpc-mspc.ca/resources/Documents/Respository/PHPC%20COVID19%20Dashboard%20Summary%2006-16-2020.pdf [accessed 2021-07-31]
39. Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CF, Askham J, et al. Consensus development methods, and their use in clinical guideline development. Health Technol Assess 1998;2(3):i-iv, 1 [FREE Full text] [Medline: 9561895]
40. Costa-Santos C, Antunes L, Souto A, Bernardes J. Assessment of disagreement: a new information-based approach. Ann Epidemiol 2010 Jul;20(7):555-561. [doi: 10.1016/j.annepidem.2010.02.011] [Medline: 20538199]
41. King N. Using templates in the thematic analysis of text. In: Essential Guide to Qualitative Methods in Organizational Research. Newbury Park, CA: Sage Publications; 2004.
42. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
43. How's My Flattening. URL: https://howsmyflattening.ca/ [accessed 2021-07-31]
44. COVID-19: Case Counts. City of Toronto. URL: https://www.toronto.ca/home/covid-19/covid-19-latest-city-of-toronto-news/covid-19-pandemic-data/covid-19-weekday-status-of-cases-data/ [accessed 2021-07-31]
45. Daily COVID-19 Dashboard. Ottawa Public Health. URL: https://www.ottawapublichealth.ca/en/reports-research-and-statistics/daily-covid19-dashboard.aspx [accessed 2021-07-31]
46. Berry I, Soucy JR, Tuite A, Fisman D, COVID-19 Canada Open Data Working Group. Open access epidemiologic data and an interactive dashboard to monitor the COVID-19 outbreak in Canada. CMAJ 2020 Apr 14;192(15):E420 [FREE Full text] [doi: 10.1503/cmaj.75262] [Medline: 32392510]
47. COVID-19 in Canada. COVID-19 Canada Open Data Working Group. URL: https://art-bd.shinyapps.io/covid19canada/ [accessed 2021-07-31]
48. Data on COVID-19 in Québec. Government of Quebec. URL: https://www.quebec.ca/en/health/health-issues/a-z/2019-coronavirus/situation-coronavirus-in-quebec [accessed 2021-07-31]
49. All Ontario: Case numbers and spread. Government of Ontario. URL: https://covid-19.ontario.ca/data [accessed 2021-07-31]
50. COVID-19 Resources. URL: https://resources-covid19canada.hub.arcgis.com/ [accessed 2021-07-31]
51. Verhulsdonck G, Shah V. Lean Data Visualization: Considering Actionable Metrics for Technical Communication. Journal of Business and Technical Communication 2020 Sep 10;35(1):57-64. [doi: 10.1177/1050651920958500]
52. Ting I, Scott N, Workman M. Charting the COVID-19 spread in Australia. ABC News. 2020. URL: https://www.abc.net.au/news/2020-03-17/coronavirus-cases-data-reveals-how-covid-19-spreads-in-australia/12060704 [accessed 2021-07-31]
53. COVID-19 Sledilnik. 2020. URL: https://covid-19.sledilnik.org/en/stats [accessed 2021-07-31]
54. Tracking the coronavirus. CBC News. 2020. URL: https://newsinteractives.cbc.ca/coronavirustracker/ [accessed 2021-07-31]
55. Nuti S, Vola F, Bonini A, Vainieri M. Making governance work in the health care sector: evidence from a 'natural experiment' in Italy. Health Econ Policy Law 2016 Jan;11(1):17-38 [FREE Full text] [doi: 10.1017/S1744133115000067] [Medline: 25819303]
56. Compton-Phillips AT. The Give a Darn Method for Outcomes Measurement, in NEJM Catalyst. NEJM Catalyst. 2018. URL: https://catalyst.nejm.org/doi/full/10.1056/CAT.18.0262 [accessed 2021-07-31]
57. Kurtzman ET, Greene J. Effective presentation of health care performance information for consumer decision making: A systematic review. Patient Educ Couns 2016 Jan;99(1):36-43. [doi: 10.1016/j.pec.2015.07.030] [Medline: 26277826]
58. A stronger and more resilient Canada: Speech from the Throne to Open the Second Session of the Forty-Third Parliament of Canada. Canada.ca. 2020. URL: https://www.canada.ca/en/privy-council/campaigns/speech-throne/2020/stronger-resilient-canada.html [accessed 2021-07-31]
59. Bernier A, Knoppers BM. Pandemics, privacy, and public health research. Can J Public Health 2020 Aug;111(4):454-457 [FREE Full text] [doi: 10.17269/s41997-020-00368-5] [Medline: 32592023]
60. Datta G, Siddiqi A, Lofters A. Transforming race-based health research in Canada. CMAJ 2021 Jan 18;193(3):E99-E100 [FREE Full text] [doi: 10.1503/cmaj.201742] [Medline: 33462147]
61. Proposed Standards for Race-Based and Indigenous Identity Data Collection and Health Reporting in Canada. Canadian Institute for Health Information. 2020. URL: https://www.cihi.ca/sites/default/files/document/proposed-standard-for-race-based-data-en.pdf [accessed 2021-07-31]
62. Disaggregated demographic data collection in British Columbia: The grandmother perspective. British Columbia's Office of the Human Rights Commissioner. 2020. URL: https://bchumanrights.ca/wp-content/uploads/BCOHRC_Sept2020_Disaggregated-Data-Report_FINAL.pdf [accessed 2021-07-31]
63. Ploem C, Suurmond J. Registering ethnicity for covid-19 research: is the law an obstacle? BMJ 2020 Sep 01;370:m3367. [doi: 10.1136/bmj.m3367] [Medline: 32873550]

64. McKenzie K. Socio-demographic data collection and equity in covid-19 in Toronto. EClinicalMedicine 2021 Apr;34:100812 [FREE Full text] [doi: 10.1016/j.eclinm.2021.100812] [Medline: 33898954]
65. COVID-19 data portal. Stats NZ. 2020. URL: https://www.stats.govt.nz/experimental/covid-19-data-portal [accessed 2021-07-31]
66. Demographic tends of COVID-19 cases and deaths in the US reported to CDC. Centers for Disease Control and Prevention. 2020. URL: https://covid.cdc.gov/covid-data-tracker/#demographics [accessed 2021-07-31]
67. Klazinga N, Stronks K, Delnoij D, Verhoeff A. Indicators without a cause. Reflections on the development and use of indicators in health care from a public health perspective. Int J Qual Health Care 2001 Dec;13(6):433-438. [doi: 10.1093/intqhc/13.6.433] [Medline: 11769744]
68. Smith PC, Mossialos E, Papanicolas I. Performance measurement for health system improvement: experiences, challenges and prospects. World Health Organization. 2008. URL: https://www.who.int/management/district/performance/PerformanceMeasurementHealthSystemImprovement2.pdf [accessed 2021-07-31]
69. Smith PC, Mossialos E, Papanicolas I, Leatherman S. Introduction. In: Performance Measurement for Health System Improvement. Cambridge, UK: Cambridge University Press; 2010:3-24.
70. Hilarion P, Suñol R, Groene O, Vallejo P, Herrera E, Saura RM. Making performance indicators work: the experience of using consensus indicators for external assessment of health and social services at regional level in Spain. Health Policy 2009 Apr;90(1):94-103. [doi: 10.1016/j.healthpol.2008.08.002] [Medline: 18829129]
71. Jakab M, Nathan NL, Pastorino G, Evetovits T, Garner S, Langins M, et al. Managing health systems on a seesaw: Balancing the delivery of essential health services whilst responding to COVID-19. World Health Organization. 2020. URL: https://apps.who.int/iris/handle/10665/336299 [accessed 2021-07-31]

Abbreviations

PT: provincial/territorial
WHO: World Health Organization

Edited by G Eysenbach; submitted 05.05.21; peer-reviewed by N Riley, A Naser; comments to author 27.05.21; revised version received 11.06.21; accepted 05.07.21; published 06.08.21.

Please cite as:
Barbazza E, Ivanković D, Wang S, Gilmore KJ, Poldrugovac M, Willmington C, Larrain N, Bos V, Allin S, Klazinga N, Kringos D
Exploring Changes to the Actionability of COVID-19 Dashboards Over the Course of 2020 in the Canadian Context: Descriptive Assessment and Expert Appraisal Study
J Med Internet Res 2021;23(8):e30200
URL: https://www.jmir.org/2021/8/e30200
doi:10.2196/30200
PMID:34280120

©Erica Barbazza, Damir Ivanković, Sophie Wang, Kendall Jamieson Gilmore, Mircha Poldrugovac, Claire Willmington, Nicolas Larrain, Véronique Bos, Sara Allin, Niek Klazinga, Dionne Kringos. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Self-Focused and Other-Focused Health Concerns as Predictors of the Uptake of Corona Contact Tracing Apps: Empirical Study

Fenne große Deters[1,2], Dr; Tabea Meier[3], Dr; Anne Milek[4], Dr; Andrea B Horn[3,5], Dr

[1]Weizenbaum Institute for the Networked Society, Berlin, Germany
[2]University of Potsdam, Potsdam, Germany
[3]University Research Priority Program, University of Zurich, Zurich, Switzerland
[4]University of Münster, Münster, Germany
[5]Center for Gerontology, University of Zurich, Zurich, Switzerland

Corresponding Author:
Fenne große Deters, Dr
Weizenbaum Institute for the Networked Society
Hardenbergstraße 32
Berlin, 10623
Germany
Phone: 49 331977 3710
Email: fenne.grosse.deters@uni-potsdam.de

Abstract

Background: Corona contact tracing apps are a novel and promising measure to reduce the spread of COVID-19. They can help to balance the need to maintain normal life and economic activities as much as possible while still avoiding exponentially growing case numbers. However, a majority of citizens need to be willing to install such an app for it to be effective. Hence, knowledge about drivers for app uptake is crucial.

Objective: This study aimed to add to our understanding of underlying psychological factors motivating app uptake. More specifically, we investigated the role of concern for one's own health and concern to unknowingly infect others.

Methods: A two-wave survey with 346 German-speaking participants from Switzerland and Germany was conducted. We measured the uptake of two decentralized contact tracing apps officially launched by governments (Corona-Warn-App, Germany; SwissCovid, Switzerland), as well as concerns regarding COVID-19 and control variables.

Results: Controlling for demographic variables and general attitudes toward the government and the pandemic, logistic regression analysis showed a significant effect of self-focused concerns (odds ratio [OR] 1.64, P=.002). Meanwhile, concern of unknowingly infecting others did not contribute significantly to the prediction of app uptake over and above concern for one's own health (OR 1.01, P=.92). Longitudinal analyses replicated this pattern and showed no support for the possibility that app uptake provokes changes in levels of concern. Testing for a curvilinear relationship, there was no evidence that "too much" concern leads to defensive reactions and reduces app uptake.

Conclusions: As one of the first studies to assess the installation of already launched corona tracing apps, this study extends our knowledge of the motivational landscape of app uptake. Based on this, practical implications for communication strategies and app design are discussed.

(J Med Internet Res 2021;23(8):e29268) doi:10.2196/29268

KEYWORDS

COVID-19; corona contact tracing app; digital proximity tracing; preventive behavior; health concern; prosocial motivation; public health; risk perception, eHealth, Corona-Warn-App; SwissCovid; contact tracing app; contact tracing

Introduction

Background

In March 2020, the World Health Organization declared the outbreak of the novel coronavirus SARS-CoV-2 a pandemic [1], and worldwide governments took radical measures to reduce the rate of transmission [2]. Most of these measures like quarantining potentially infected individuals, reducing social contact, and wearing face masks have been used for centuries to limit the spread of contagious diseases [3,4]. Technological advances in the 21st century have added contact tracing apps (CTAs) to our toolbox. Unsurprisingly, many governments are interested in CTAs as an additional measure to keep the pandemic under control [5]. In December 2020, a database by the MIT Technology Review listed 48 countries that have or are in the process of developing CTAs [6]. In the overwhelming majority of cases, the usage of such an app is voluntary or is planned to be voluntary. Simulation studies suggest that 56% of a population must use the app for an effect on the overall development of case numbers [7]. To reach this goal, it is important to understand what motivates citizens to adopt CTAs [8]. CTAs are a new phenomenon and, so far, early research mostly assessed app uptake intention before CTAs were launched [9]. Moreover, only a few studies have focused on underlying psychological factors motivating app uptake [8,10-12]. This study aimed to add to this line of research by investigating the role of self-focused and other-focused concerns regarding COVID-19 as so far understudied predictors of the adoption of two already launched CTAs.

COVID-19 and Contact Tracing

An important measure to reduce the transmission of COVID-19 is contact tracing [13]. Individuals who have been physically close to an infected person receive a warning by health officials that they might have caught the disease and, even if they are not (yet) showing any symptoms, might be a spreader of the virus [14]. Warned individuals should then self-isolate and get tested [15,16]. By breaking the chain of infection, contact tracing can reduce the spread of the virus and help to contain the pandemic [13].

However, manual contact tracing has three problems. First, with exponentially rising numbers of cases, health officials are quickly overwhelmed with the workload. This results in a slower pace or even complete failure of informing individuals. Second, infected individuals might not recall all encounters they had in the critical time period. Finally, in case of contact with strangers like on public transportation, they might simply lack information on individuals with whom they have spent time in close proximity [5].

CTAs can mitigate these problems. Automatic contact tracing reduces the time between a positive test result and sending a warning to contacts of the infected person [17]. This is crucial for breaking the chain of infection, especially since presymptomatic transmission of COVID-19 appears to be common [14]. Automatic tracing via apps can also be scaled up more easily [18] and CTAs do not rely on the information on contacts provided by the infected individual [15].

Corona CTAs

CTAs are apps for smartphones that keep track of other smartphones that have been in close physical proximity, and in case a smartphone owner tests positive for COVID-19, they allow sending a warning to these other smartphones. CTAs can be broadly classified into three different categories based on the role of the central server and the types of data that are stored by it (centralized, decentralized, and hybrid) [5]. These different architectures have implications for privacy protection and data security. Hence, they might influence the outcome of benefit-risk analyses and might determine the willingness to adopt a technology [19-21].

This study focuses on two decentralized CTAs officially launched by governments (Corona-Warn-App, Germany; SwissCovid, Switzerland). Decentralized CTAs prioritize privacy protection. The users' identities are unknown to both other users and the central server [5]. Both the German and Swiss CTAs rely on the Google/Apple API, use Bluetooth Low Energy Technology to communicate with other smartphones, and can be downloaded for free [22-24]. SwissCovid (Switzerland) was released on June 25, 2020, to the general public and was downloaded 3 million times until March 2021 [25]. Corona-Warn-App (Germany) was released on June 16, 2020, and was downloaded 26.5 million times until March 2021 [26]. With around 6.8 million smartphone users in Switzerland [27] and around 66.5 million smartphone users in Germany [28], only 44% and 40%, respectively, of all potential users, in the most optimistic scenario, might have already adopted the apps. This highlights the need to understand drivers to increase app uptake.

Benefits of App Uptake for Oneself and For Others as Potential Motivators

The benefits of using CTAs for oneself are not straightforward. Other apps in the health care sector like fitness or therapy apps promise to address a personal health problem or lower a risk for the user [12]. However, CTAs do not protect the user from COVID-19. These apps only warn individuals *retrospectively* that they have been exposed to the virus after a contact person has tested positive [5]. The value provided by CTAs results in a social dilemma–like situation. On the one hand, a wide acceptance of CTAs benefits all because this could lower the spread of the virus in the population. On the other hand, each individual has costs (time, inconvenience, etc) in case of installing the app, but no direct and clear health benefits or risk reduction [29].

Users only gain information on the likelihood of currently being infected. However, while not the focus of CTAs, individuals might use this information to evaluate their past behaviors in order to protect their health in the future. If warnings occur, they can review situations they have been in during the critical time period from which the warning resulted and try to avoid these.

Another direct benefit for a user is the decreased risk of unknowingly infecting others. Even if individuals feel healthy, they might be carriers of COVID-19 and spread the virus with potentially severe consequences for others [14]. Whether visiting a family member or spending time in a restaurant, a multitude of situations involve the risk of not only getting infected, but also accidentally harming others and having to face the guilt [30]. Particularly for individuals who expect to experience no dramatic symptoms if they catch COVID-19, reducing the risk of unknowingly infecting others might be a major selling point for CTAs. In short, app uptake might be motivated by being concerned about others rather than by being concerned about one's own health. Our study examines the role of self-focused and other-focused concerns about COVID-19 for app uptake.

Concerns as Drivers of Health-Compliant Behaviors

From the broader literature on preventive behaviors in the health context, it is evident that risk perception influences the likelihood of taking protective measures. The Protection Motivation Theory [31,32] and the Health Belief Model [33,34], two prominent theories in the area of health behavior promotion, suggest that risk perception can be a strong driver for individuals to engage in preventive behaviors [35]. Several studies have found positive associations between risk perception and health-compliant behaviors like social distancing and hand washing during the current pandemic [36,37] and previous disease outbreaks [35]. In other words, adoption of preventive behaviors is more likely if individuals think that they might be individually affected by the health problem. Risk perception can be approached from a cognitive or an emotional perspective. The emotional component of risk perception is characterized by worrying or being concerned about a threat. In the literature, these emotional facets of risk perceptions are seen as important predictors of favorable health behaviors [30]. In a recent study testing several theoretically derived variables as predictors of the intention to install a CTA, feeling anxious that oneself or a close other contracts COVID-19 emerged as a significant predictor [11]. There are some contradicting results in the literature, however, which have been interpreted by some as reflecting the effect of fear control instead of danger control responses. Too strong emotional reactions to a risk might overwhelm the individual, particularly if measures to reduce the threat are perceived as only moderately effective, and lead to defensive reactions and the rejection of preventive behaviors [38]. This boomerang effects can be tested by including curvilinear associations in the analysis [38-40].

Overall, it can be argued that certain characteristics of the situation during a novel pandemic likely impede with cognitive assessment of risks and render emotional aspects more important. The situation is quickly evolving, high levels of uncertainty exist due to often preliminary or contradictory information, and there is an acute threat [30]. Research conducted during the influenza A/H1N1 pandemic showed that emotional aspects of risk perception predicted protective behaviors over and above cognitive aspects [41] and emerged as stronger and more consistent predictors [42]. Hence, this study focused on concerns as the emotional dimension of risk perception.

Concern for One's Own Health and App Uptake

Feeling at risk might motivate individuals to engage in preventive behaviors. However, according to the Protection Motivation Theory [31,32] and the Health Belief Model [33,34], this should only be the case if a certain behavior is perceived as effective in reducing the threat [35]. As outlined above, CTAs are not designed to decrease the risk of COVID-19 for an individual user. This raises the question whether concern for one's own health is linked to app installation.

Several studies have reported that many participants (32% to 84%) list "protect my own health" among their reasons for installing such a CTA [18,43-47]. The association between perceived personal risk and intended app uptake was only assessed by a few studies and with mixed results. In some studies, the cognitive dimension of risk perception, namely the severity and susceptibility of infection, did not significantly predict the intention to install a CTA [8,10], whereas other studies found positive associations between personal risk and willingness to adopt a CTA [16,18,48-50]. These inconclusive results and the lack of studies assessing the emotional dimension instead of the cognitive dimension of risk perception warrant further research.

Another open question is whether very high levels of concern are linked to lower instead of higher app uptake. Too intense concerns might be overwhelming and may paralyze individuals or provoke defensive reactions, and hence, hinder them to take preventive actions [30]. Such a curvilinear relationship between concern and health-related behaviors has not yet been tested in respect to CTAs.

Concern for Others and App Uptake

Infectious diseases like COVID-19 that are transmitted by close contact to others inherently have a social dimension [30]. As social animals, we worry about not only our own welfare but also the welfare of others [51]. In the context of COVID-19, research has shown that consideration of others plays a role in the adoption of preventative health behaviors like wearing face masks, curtailing social contact, being willing to get vaccinated, or refraining from concealing potential COVID-19 symptoms [52-54]. Given that CTAs are intended to reduce the risk of the spread of the virus and hence protect others, it seems reasonable to assume that concern to infect others is associated with app uptake.

Previous research has shown that potential users are aware that CTAs protect others, and for a majority (52%-68%), this is an important reason to install the app [18,43,45-47]. In a discrete choice experiment on preferences for different app attributes, positive societal effects had a large impact on the probability to install a CTA [50]. Participants preferred app configurations that promised the prevention of deaths and long-term financial problems of households. However, as the authors note themselves, despite the collective framing, participants might have had their own well-being in mind, when evaluating the different app configurations. Another experimental study directly compared different motivations for app installation [12]. Participants either learned that by using the app they could (1) make an important contribution to their own health, (2) make an important contribution to the health of the population, or (3) both. Participants' intentions to install a CTA were the highest if advertised with a procommunal effect and the lowest if only advertised with a personal health benefit. Apart from the three benefit appeals, two privacy designs and two convenience designs were manipulated in a full factorial experimental design. The authors do not report any tests for interaction effects. This warrants a cautious interpretation because the main effects could be qualified by an interaction. Moreover, the study assessed intention to use instead of actual app uptake. Ample research has demonstrated that intentions do not necessarily translate to actual behavior [55]. Accordingly, two studies on CTAs reported a gap between the intention and the actual installation of the app [45,56]. Moreover, due to the clear prosocial framing in the experiment, social desirability bias might have been an important factor influencing its outcome. Rejecting the installation of an app that benefits oneself is certainly more socially acceptable than being indifferent to the death and suffering of others [57].

Taken together, it remains an open question how self-focused and other-focused concerns relate to app installation. This might be relevant information to tailor promotion strategies, particularly in light of the unusual incentive structure CTAs offer.

Our Study

This study explored the association between self-focused and other-focused concerns and app uptake. Given the novelty of the situation, this research is exploratory and assesses the following research questions: (1) What is the role of concern for one’s own health and concern to infect others in predicting the installation of a CTA? (2) Are very high levels of concern in comparison to moderate levels linked to a lower instead of higher likelihood of using a CTA?

In order to obtain robust results, we included potential confounders in our analyses. Apart from demographics like gender and age [18,49], disagreement with the classification of COVID-19 as a serious health crisis and the evaluation of the government’s pandemic policies are of interest. Both have been linked to the likelihood of taking preventive measures during pandemics [18,48,58-60]. Moreover, individuals holding the believe that an infectious disease has been hyped up can be expected to be less concerned about its effects. Moreover, feeling that the government is handling the health crisis satisfactorily might be linked to concerns. A positive evaluation of the government’s pandemic policies might be reassuring and associated with lower levels of risk perception [61]. On the contrary, perceiving the drastic measures like lockdowns issued by the governments of Switzerland and Germany [62] as justified might be associated with higher corona-related concerns. Moreover, a study in Singapore showed that higher confidence that the government would be able to handle the pandemic predicted installation of a CTA [63]. Hence, we included these variables in our study to control for confounding effects.

As another robustness check, we tested whether the association between app uptake and concerns depends on the timing of measuring concerns. Two opposing processes might influence concurrent associations between concerns and app uptake. First, engaging in recommended health behaviors might reduce a perceived health threat [64]. Hence, assessing concerns and app uptake only at the same point in time might result in finding a negative association or no association even if concerns actually motivated installation. Second, instead of feeling protected by the app, some participants expected increased feelings of anxiety as a result of app usage [18,47]. App users potentially receive warnings that they have been in close proximity to a COVID-19 carrier, and the icon of the app on the smartphone might act as a reminder of a threat. Hence, associations between concerns and app uptake in cross-sectional studies might reflect the reverse temporal order, that is, app usage preceding high levels of concern. Therefore, we included longitudinal analyses predicting app uptake with concerns measured before app release.

So far, research on drivers for the adoption of CTAs is scarce. Exploring the motivational landscape might enhance communication strategies and create ideas for features in the app. This study examines the interplay between self-focused and other-focused concerns in predicting app uptake while considering potential confounders, considering temporal aspects, and testing a curvilinear association. In contrast to most previous research, we focused on the emotional aspect of risk perception and assessed not just the intention to install a CTA but predict the actual (self-reported) installation of two already launched apps in real life.

Methods

Sample and Procedures

Data were collected online on the survey-platform Qualtrics [65] in two waves (T1: April 16-June 27, 2020; T2: July 1-September 18, 2020) as part of a multinational initiative studying how our everyday lives are affected by COVID-19 (Ashokkumar & Pennebaker, The Pandemic Project: Exploring the Social Dynamics of COVID-19, unpublished, 2020). The subsample for this study consists of German-speaking participants from Switzerland and Germany, which are two countries that had launched a CTA shortly before the data collection of the second wave took place. Reported COVID-19 case numbers and deaths remained low in both countries after the introduction of the CTA and the beginning of the data collection until its end. Similarly, measures and policies to control COVID-19 were comparable in both countries and remained largely unchanged during this time [66-68]. The link to the study was distributed by all authors via social media, mailing lists, newsletters, the Senior Citizens’ University of Zurich study participant pools, and websites of universities, as well as the website of the popular science magazine “Psychologie heute.” Fifteen participants were recruited via the participant recruitment platform Prolific. Prolific participants were eligible if German was their first language, they were living in Germany or Switzerland, and they were at least 50 years old. The goal here was to diversify the sample by recruiting more middle-aged and older participants.

With their permission, participants of the first wave received an invitation for the second wave, which included an ID to link the data of both waves. Additionally, new participants were recruited for the second wave via the aforementioned recruitment channels. The app-related questions were only included in the second wave after the launch of the CTAs. The final sample for the main analyses (N=346) consisted of all participants who answered those questions in wave 2. The sample for the additional longitudinal analyses included a subsample (N=270) of participants who had also participated in the first wave. Prior to data analyses, all data were anonymized.

Table 1 provides an overview of relevant demographic variables. In comparison to the general population in Switzerland and Germany, the sample included more participants who identified as female, had a higher education, and indicated their political orientation as left [69-74].

After opening the link to the study, general information about the topic of the study was provided. Participants learned that their participation was voluntary, and received information about data protection and their rights. After giving informed consent, participants provided demographic information. Next, participants were prompted to write about their thoughts and feelings related to the coronavirus outbreak. Then, participants proceeded to answer questions related to COVID-19 and their daily life during the pandemic (waves 1 and 2) and questions about the installation of a CTA (wave 2). In the last step, participants received feedback (eg, how preoccupied they are with the pandemic) and advice based on psychological research on how to cope with the situation. Prolific participants were compensated with £2.5 (US $1.4), and psychology students could receive course credit if eligible.

Table 1. Demographics (N=346).

Demographic variable	Value, n (%) or mean
Subsample	
Germany	114 (32.9%)
Switzerland	232 (67.0%)
Gender	
Male	82 (23.7%)
Female	262 (75.7%)
Diverse/no answer	2 (0.6%)
Age (years)	
18-29	70 (20.2%)
30-39	74 (21.4%)
40-49	66 (19.1%)
50-59	32 (9.2%)
60-69	46 (13.3%)
70-79	50 (14.5%)
80-89	8 (2.3%)
Mean age (years)	46.65
Highest education	
Higher education (bachelor’s degree, master’s degree, PhD)	232 (67.1%)
Higher education entrance qualification	59 (17.1%)
Vocational training	45 (13.0%)
Lower to intermediate secondary education	8 (2.3%)
Other/no degree	2 (0.6%)
Political orientation	
Extremely or somewhat left wing	219 (63.3%)
In the middle	63 (18.2%)
Extremely or somewhat right wing	32 (9.2%)
I do not want to tell	32 (9.2%)

Measures

Predictors

Table 2 and Table 3 provide information on the measures included as nondemographic predictors. All constructs were measured with a single item on a 5-point Likert scale ranging from “Not at all” (score 1) to “To a great deal” (score 5). “Concern self” and “Concern others” were measured in both wave 1 (T1) and wave 2 (T2).

Table 2. Nondemographic predictors.

Variable	Item English	Item German	Score, mean (SD)
1. Concern self (T2)	To what degree are you worried about getting COVID-19.	In welchem Ausmaß machen Sie sich Sorgen, selbst an COVID-19 zu erkranken.	2.15 (0.92)
2. Concern self (T1)	To what degree are you worried about getting COVID-19.	In welchem Ausmaß machen Sie sich Sorgen, selbst an COVID-19 zu erkranken.	2.10 (1.00)
3. Concern others (T2)	To what degree are you worried about unknowingly infecting others.	In welchem Ausmaß machen Sie sich Sorgen, unwissentlich Andere zu infizieren.	2.93 (1.22)
4. Concern others (T1)	To what degree are you worried about unknowingly infecting others.	In welchem Ausmaß machen Sie sich Sorgen, unwissentlich Andere zu infizieren.	3.11 (1.20)
5. Satisfaction with the government (T2)	I am satisfied with how my government has handled the COVID crisis.	Ich bin damit zufrieden wie meine Regierung mit der COVID-19 Krise umgegangen ist.	3.70 (0.86)
6. Not perceiving COVID-19 as a health crisis (T2)	To what degree do you feel that people are making too big a deal about COVID-19.	Inwieweit sind Sie der Meinung, dass die Leute eine zu große Sache aus COVID-19 machen.	2.11 (1.11)

Table 3. Correlations of nondemographic predictors.

Variable	Concern self (T2)	Concern self (T1)	Concern others (T2)	Concern others (T1)	Satisfaction with the government (T2)	Not perceiving COVID-19 as a health crisis (T2)
Concern self (T2)						
r	1	0.62	0.32	0.24	0.18	−0.34
P value	–[a]	<.001	<.001	<.001	.001	<.001
Concern self (T1)						
r	0.62	1	0.15	0.23	0.16	−0.28
P value	<.001	–	.01	<.001	.01	<.001
Concern others (T2)						
r	0.32	0.15	1	0.66	0.13	−0.29
P value	<.001	.01	–	<.001	.02	<.001
Concern others (T1)						
r	0.24	0.23	0.66	1	0.13	−0.34
P value	<.001	<.001	<.001	–	.03	<.001
Satisfaction with the government (T2)						
r	0.18	0.16	0.13	0.13	1	−0.38
P value	.001	.01	.02	.03	–	<.001
Not perceiving COVID-19 as a health crisis (T2)						
r	−0.34	−0.28	−0.29	−0.34	−0.38	1
P value	<.001	<.001	<.001	<.001	<.001	–

[a] N/A: not applicable.

Outcome

The outcome was measured in wave 2 after the release of the CTAs in Switzerland and Germany. Participants were asked whether they have installed a CTA like “SwissCovid” or “Corona-Warn-App” with the following response options: “Yes;” “Yes, but already uninstalled;” “No;” and “No, but I will likely do it.” The majority of the participants (202/346, 58.4%) had installed a CTA. The remaining participants indicated that they had not installed the app (111/346, 32.1%), had not yet installed it but will likely do so (25/346, 7.2%), or had already uninstalled it (8/346, 2.3%).

Analytical Strategy

We recoded answers on app installation into a binary variable. If participants answered with "yes," their answer was recoded as "app currently installed," and all other answers ("no;" "not yet installed it but will likely do so;" and "yes, but already uninstalled it") were recoded as "app currently not installed." Using the R package "stats" [75], we ran logistic regression models to predict app installation. First, two separate models for the predictors "Concern self (T2)" and "Concern others (T2)" were calculated to show their univariate effect (models M1a and M1b). Next, both variables were included in the same model simultaneously to assess their unique contribution (model M2). Finally, we explored how the results changed when controlling for "Satisfaction with the government" and "Not perceiving COVID-19 as a serious health crisis." Moreover, we added nationality, age, gender, highest education, and political orientation as control variables (model M3). To ease interpretation, we transformed the logistic regression coefficients to odds ratios (ORs), that is, the expected change in the odds of having the app installed if the predictor increases by one unit or changes from the reference category to another category in the case of categorical variables [76].

At each step, likelihood ratio tests showed a significantly improved model fit of the model with more predictors in comparison to the previous model with fewer predictors [77]. Starting with a model with the control variables and then adding "Concern self" and "Concern others" also resulted in an improved model fit. Because of the focus of the paper, we began with the two concern variables. Participants with missing values for any of the variables were excluded, which resulted in a sample size of 340 for these analyses. To get unbiased regression coefficients, an events per variable ratio of 1:10 or higher has been recommended for logistic regression [78,79]. With 140 participants in the category with fewer answers ("app currently not installed"), the events per variable ratio in the final model was 1:10.

Model Evaluation

As recommended for logistic regression models, we compared observed and predicted values to evaluate the fit of our model [80]. We calculated the area under the receiver operating characteristic curve (AUC, also known as c statistic). The AUC is the proportion of randomly drawn pairs of participants with different observed outcomes for which the model correctly predicts a higher probability of app uptake for the participant who has the app versus for the participant who does not. It ranges from 0.5 to 1. A value of 0.5 indicates that the model is not better than completely random assignment, while a value of 1 shows perfect performance [77].

Longitudinal Analyses

Concerns were measured at T1 as well before the release of the CTAs in Germany and Switzerland. This allowed us to test whether the association between app uptake and concerns depends on the timing of measuring concerns. We repeated all analyses with the concern variables measured at T1 (models M4a-M6). Not all participants had completed the questionnaire at T1, and analyses could only be performed with a sample of 270. The proportion of app users remained the same in this sample. Moreover, we assessed whether changes in concerns were associated with app uptake. We ran linear regression models predicting "Concern self (T2)" with app uptake while controlling for "Concern self (T1)" and our control variables. We repeated this analysis for "Concern others (T2)."

Curvilinear Association

To test whether too high levels of concern lead to defensive reactions, a curvilinear association of concerns with app uptake was tested. We mean-centered "Concern self (T2)," calculated the squared term for "Concern self (T2)," and added it to the final model. We repeated the same steps for "Concern others (T2)" (models M7 and M8).

Results

Predicting App Uptake With Concerns

All results are displayed in Table 4. Both "Concern self (T2)" and "Concern others (T2)" showed significant univariate associations with app uptake. Namely, the higher the concern, the higher the likelihood of currently using a CTA (models M1a and M1b). However, if both concerns were included as predictors, only "Concern self (T2)" significantly predicted app installation with an OR of 1.73. "Concern others (T2)" did not contribute significantly to the prediction of app uptake over and above worry for one's own health (model M2).

Controlling for demographic variables and attitudes toward the government and the pandemic in general reduced the effect of "Concern self (T2)" slightly (model M3). Holding all other variables constant, with a one unit increase in "Concern self (T2)," the odds of currently having an app installed were 1.64 times higher. As expected, the less individuals perceived COVID-19 as a serious health crisis, the lower was the probability of app installation. Higher "Satisfaction with the government" was significantly associated with increased app uptake. Three demographic variables emerged as significant predictors. First, all else equal, participants in the Swiss subsample were more likely to have the app installed than participants in the German subsample. It should be noted that because the Swiss and German subsamples were not representative of the respective populations, the significant effect of the subsample in our study should not be interpreted as differences in app uptake between the two countries in general. Moreover, access to study populations and available recruiting strategies differed slightly between the authors in Switzerland and Germany. The variable was included in the model to assess the influence of concerns independent of these differences. Second, with older age, the likelihood of app uptake decreased. Finally, in comparison with the reference category of individuals with a degree in higher education, individuals with a "Higher education entrance qualification" were less likely to have a CTA installed.

Table 4. Logistic regression models predicting app uptake.

Model	*b*	SE	*P* value	95% CI		OR[a]
				Lower	Upper	
M1a						
Concern self (T2)	0.58	0.13	<.001	0.32	0.84	1.78
M1b						
Concern others (T2)	0.19	0.09	.03	0.02	0.37	1.21
M2						
Concern self (T2)	0.55	0.14	<.001	0.28	0.83	1.73
Concern others (T2)	0.07	0.10	.46	−0.11	0.26	1.07
M3						
Concern self (T2)	0.50	0.16	.002	0.19	0.81	1.64
Concern others (T2)	0.01	0.11	.92	−0.21	0.24	1.01
Satisfaction with the government	0.45	0.17	.007	0.13	0.79	1.57
Not perceiving COVID-19 as a health crisis	−0.35	0.13	.007	−0.61	−0.10	0.70
Subsample Switzerland	0.65	0.29	.02	0.09	1.22	1.91
Gender female	−0.54	0.33	.10	−1.19	0.09	0.58
Age	−0.03	0.01	.002	−0.04	−0.01	0.98
Education (reference: higher education)						
Higher education entrance qualification	−1.14	0.36	<.001	−1.85	−0.44	0.32
Vocational training	0.39	0.40	.33	−0.38	1.21	1.48
Lower to intermediate secondary education	0.15	0.85	.86	−1.47	1.96	1.17
Other/no degree	−0.64	1.49	.67	−3.96	2.69	0.53
Political orientation (reference: in the middle)						
Extremely or somewhat left wing	0.05	0.34	.87	−0.62	0.71	1.06
Extremely or somewhat right wing	−0.80	0.51	.12	−1.82	0.20	0.45
I do not want to tell	−0.56	0.52	.28	−1.59	0.45	0.57

[a] OR: odds ratio.

Model Evaluation

For the final model with all variables, the AUC was 0.74. Hence, the model correctly predicted for 74% of all nonapp user/app user pairs a higher probability for a participant who indeed has the app installed.

Longitudinal Analyses

Repeating the analyses with concerns measured at T1 before app release, the pattern of results remained the same (Multimedia Appendix 1). “Concern self (T1)” significantly predicted higher app uptake with a similar OR (OR_{T1} 1.81, OR_{T2} 1.64), whereas “Concern others (T1)” had no significant effect. The AUC was 0.79. Accordingly, the linear regression models predicting change in concerns revealed no significant effect of app uptake on change in “Concern self” (b=0.14, t_{255}=1.44, P=.15) or change in “Concern others” (b=−0.00, t_{255}=−0.00, P>.99) (Multimedia Appendix 2 and Multimedia Appendix 3).

Curvilinear Association

In the last step (models M7 and M8; Multimedia Appendix 4 and Multimedia Appendix 5), we tested for a curvilinear association between concerns and app uptake. For the quadratic term of “Concern self (T2),” no significant effect emerged (b=−0.09, SE=0.12, P=.44, 95% CI −0.32 to 0.15, OR 0.91). Similarly, the quadratic term in the model for “Concern others (T2)” was also nonsignificant (b=−0.08, SE=0.08, P=.32, 95% CI −0.25 to 0.08, OR 0.92). The likelihood ratio test confirmed that including curvilinear effects of concerns did not significantly improve model fit.

Discussion

Summary and Discussion of the Findings

As one of the first studies to assess the installation of already launched CTAs, this study contributes to our understanding of different motivations for app uptake. The results showed that concern for one’s own health predicts the installation of a decentralized CTA (OR 1.64, P=.002). Meanwhile, the concern to infect others did not contribute significantly to the prediction of app uptake over and above self-focused concern (OR 1.01, *P*=.92). In other words, individuals who had higher levels of worry to infect themselves with COVID-19 had downloaded a CTA with a higher probability, while being more or less concerned about unknowingly infecting others did not make a difference. This pattern held while controlling for demographics and attitudes toward the government and the pandemic. Longitudinal analyses replicated these results, thus supporting their robustness, and indicated that app uptake was not linked to changes in concerns. No evidence was found that “too much” concern leads to defensive reactions and reduces app uptake.

CTAs provide the following prominent and direct benefit to each individual user: more information on their likelihood to be currently infected with COVID-19. A warning by the app can prevent an individual from infecting someone else if appropriate measures like self-isolation are taken. Hence, it is surprising that in comparison to being concerned about one’s own health, being concerned about unknowingly infecting others does not significantly predict app uptake, particularly as participants in our sample reported on average more concern for others than for themselves and the variability of this variable was higher. Previous studies have shown that individuals are aware of the potential of the app to protect others [18,43,45-47]. However, perhaps concern for others is just not a sufficiently strong driver to motivate individuals to overcome the hassles and potential disadvantages of app installation. Research assessing both self-focused and other-focused concerns in the context of transmission-mitigating behaviors is scarce. Guillon and Kergall [48] found a similar pattern in their study. Participants who expected high individual health consequences were significantly more likely to be willing to use a CTA in the future. However, expected health impacts of COVID-19 in their country of residence did not predict the intention to use such an app.

Our result that self-focused concern is associated with app uptake is in line with studies that assessed the role of personal threat in predicting the intention to install a CTA [16,18,49,50] and other COVID-related health behaviors [36,37]. Given that CTAs are not designed to decrease the risk of a COVID-19 infection for an individual user, these results are nevertheless unexpected. It is possible that individuals are not aware of the limitations of CTAs in that regard. Maybe they falsely assume that using the app will protect them from infection. In a study on the Australian CTA, the majority of the participants thought that the app would detect when COVID-19 carriers are near them [81]. In a qualitative study conducted in Germany and Switzerland before the release of the CTAs, the same misconception was expressed by several participants [82]. Some of the official communication around the assessed CTAs might foster such a misunderstanding. For example, in a promotion video for the German CTA, the app promises “Ich beschütze Dich und sage Bescheid, wenn es Ernst wird” (I will protect you and I will let you know when it gets serious) [83].

Even if individuals are aware of how CTAs work, they might perceive the apps as at least somewhat effective in reducing the threat. As already outlined in the Introduction section, individuals might use the information provided by the app to evaluate their past behavior and avoid certain situations in the future.

It is also possible that the association between self-focused concern and app uptake is not due to an expected health benefit, but rather due to a higher need for uncertainty reduction, which is linked to higher anxiety [84]. Maybe getting more information on the current likelihood of being infected, as imperfect as the information may be, might be perceived as positive. Individuals might hope that elevated levels of concern due to situations with contact to many others like grocery shopping will be mitigated by monitoring potential warnings by the app. Another explanation is that individuals with high concerns regarding their personal health are just in general more motivated to follow any measures that promise to control the crisis, despite no direct personal health benefit.

Starting to use a CTA might provoke emotional responses. However, our longitudinal analyses do not support the possibility that app uptake is linked to changes in levels of concern. We found no evidence for increases in concern due to an increased awareness of the threat through app usage or decreases due to having adopted a recommended measure. At the very least, such effects did not appear uniformly across participants. This is in line with previous studies, which also did not report a significant change in concerns depending on app uptake [56,85]. During times with higher numbers of COVID-19 cases and hence a higher prevalence of warnings by the app, this might change.

Practical Implications

While more research is needed to solidify, explain, and test the generalizability of our results, they suggest some practical implications. First, as self-focused concern could be the underlying motivation to install a CTA, implementing features that provide more information to users about the time of the exposure to COVID-19 might increase app uptake. This information might be perceived as helping individual risk management. With more detailed information on the timing of exposure, app users might be able to learn which situations hold a high risk of exposure and hence can try to avoid them in the future (eg, grocery shopping on a Saturday versus during the week). Moreover, they might remember more details about the specific situation (eg, whether it was outside or inside or whether everybody was wearing masks) and therefore feel enabled to better evaluate the actual risk of an infection. Decentralized CTAs already store timestamps. These data are necessary to calculate whether an encounter was long enough to pose a substantial risk and for how long a warning needs to be displayed until the app user’s potential period of infectiousness ends [5]. Naturally, there is always a trade-off between data protection and the benefits that come with collecting and sharing more data [21]. If warnings include detailed information on the timing of encountering a COVID-19 carrier, the identity of the infected person might get exposed. This could be mitigated if the granularity of the time information adapts in a way that, for example, at least five encounters with noninfected users have taken place during the same time window as well.

Second, app promotion could build on the fact that if individuals are worried about a COVID-19 infection, it is in their own best interest that everybody around them, like family, friends, or colleagues, uses a CTA. Highlighting this in official campaigns and framing messages in a way that directly encourages individuals to ask their close contacts to install the app might be an effective way to increase app uptake. After all, it is probably easier to ignore an official advertisement campaign for prosocial behavior than to ignore grandma, friends, or colleagues when they nag, beg, or demand us to do something in their favor. Moreover, reciprocity norms will likely ensure that both parties engaged in that conversation end up installing a CTA. In contrast to other health-compliant behaviors like wearing masks, app usage is not directly visible. Encouraging discussions about app usage might intensify normative social influence and hence increase the socially desirable installation of a CTA [8,11,30].

Limitations

Several limitations of our study need to be considered when drawing conclusions. First, we used nonprobability sampling. However, as recommended in this case, we aimed for a broad sample, using different recruitment options and a diverse set of incentives for participation (money, course credit, personal feedback, and possibility for self-reflection). Moreover, our research question was neither the main focus of the questionnaire nor mentioned during recruitment, which decreased the risk of an association between self-selection and the target outcome [86]. Nevertheless, it is important to note that we used a web survey, and hence, participants might have a higher affinity for technology use, and that in respect to demographics, our sample is not representative of the general population in Switzerland and Germany. Therefore, we included demographic variables in our regression models to assess the association of concerns with app uptake independent of potential differences between demographic groups. The results should be carefully interpreted in consideration of the specific nature of our sample. Generalizability is also limited by the fact that our study only assessed the voluntary installation of two decentralized CTAs in two Western industrialized countries at a specific stage of the pandemic (after the first but before the second wave of COVID-19) [67,68] and in the context of moderate COVID-19–related policy responses [66]. It remains unclear whether the results hold for different app types and in different contexts [9].

Second, due to the nonexperimental nature of our data, it cannot provide evidence for causality. However, the experimental manipulation of health concerns during an ongoing pandemic would be ethically at least questionable. Moreover, the nonexperimental nature of our data collection facilitated the assessment of app installation instead of the intention for app uptake, avoiding potentially misleading conclusions. We strengthened our results by controlling for relevant variables that might be confounders. Our analyses showed that the association of self-focused concern with app uptake is not driven by the attitude toward COVID-19 or the government. When testing for reverse temporal order, we found that increased self-focused concern precedes app uptake and might hence be a driver of app installation. Nevertheless, the results on the predictors of app uptake only allow for the most cautious causal interpretation and should be considered only as hints toward such a relationship [87].

Third, we measured our predictor variables with single items. Particularly our item inquiring about participants' concerns of unknowingly infecting others does not allow us to differentiate between concerns to infect close ties like partners or friends and concerns to infect colleagues at work or even complete strangers. While ample research shows that despite widespread criticism, single-item measures are not necessarily problematic in respect to their psychometric properties [88,89], future research would certainly benefit from using multi-item measures that allow for more differentiated insights.

Conclusion

This study adds to the growing body of early research on CTAs. Hopefully, once the majority of the world's population will have been vaccinated, the pandemic spread of COVID-19 will end and the use of CTAs will not be necessary anymore. However, diseases that quickly spread in the population have always been a threat and will likely continue to be [90]. If anything, increased mobility [91] and anthropogenic pressure on the environment [92] will make a new pandemic more likely. In case of a similar outbreak, reactivating privacy-preserving CTAs might help us to be better equipped to quickly contain new diseases, while reducing disruptions of normal life. Hence, a deeper understanding of individuals' motivations to install CTAs is important not only right now but also in the future.

Acknowledgments

This work was supported by the Federal Ministry of Education and Research of Germany (BMBF) [grant number 16DII127 ("Deutsches Internet-Institut")]; the Swiss National Science Foundation [grant number SNF 31CA30_196255]; and the Jacobs Foundation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Logistic regression models predicting app uptake with concerns (T1).
[DOCX File , 20 KB - jmir_v23i8e29268_app1.docx]

Multimedia Appendix 2

Linear regression model predicting the change in “Concern self” with app uptake.
[DOCX File , 18 KB - jmir_v23i8e29268_app2.docx]

Multimedia Appendix 3
Linear regression model predicting the change in “Concern others” with app uptake.
[DOCX File , 18 KB - jmir_v23i8e29268_app3.docx]

Multimedia Appendix 4
Logistic regression model (M7) examining a curvilinear association between “Concern self” (T2) and app uptake.
[DOCX File , 19 KB - jmir_v23i8e29268_app4.docx]

Multimedia Appendix 5
Logistic regression model (M8) examining a curvilinear association between “Concern others” (T2) and app uptake.
[DOCX File , 19 KB - jmir_v23i8e29268_app5.docx]

References

1. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. World Health Organization. URL: https://tinyurl.com/6rty5cdd [accessed 2021-03-29]
2. Nicola M, Alsafi Z, Sohrabi C, Kerwan A, Al-Jabir A, Iosifidis C, et al. The socio-economic implications of the coronavirus pandemic (COVID-19): A review. Int J Surg 2020 Jun;78:185-193 [FREE Full text] [doi: 10.1016/j.ijsu.2020.04.018] [Medline: 32305533]
3. Gensini GF, Yacoub MH, Conti AA. The concept of quarantine in history: from plague to SARS. J Infect 2004 Nov;49(4):257-261 [FREE Full text] [doi: 10.1016/j.jinf.2004.03.002] [Medline: 15474622]
4. Matuschek C, Moll F, Fangerau H, Fischer JC, Zänker K, van Griensven M, et al. The history and value of face masks. Eur J Med Res 2020 Jun 23;25(1):23 [FREE Full text] [doi: 10.1186/s40001-020-00423-4] [Medline: 32576263]
5. Ahmed N, Michelin RA, Xue W, Ruj S, Malaney R, Kanhere SS, et al. A Survey of COVID-19 Contact Tracing Apps. IEEE Access 2020;8:134577-134601. [doi: 10.1109/access.2020.3010226]
6. O'Neill P, Ryan-Mosley T, Johnson B. A flood of coronavirus apps are tracking us. Now it’s time to keep track of them. MIT Technology Review. 2020. URL: https://www.technologyreview.com/2020/05/07/1000961/launching-mittr-covid-tracing-tracker/ [accessed 2021-01-25]
7. Hinch R, Probert W, Nurtay A, Kendall M, Wymant C, Hall M, et al. Effective Configurations of a Digital Contact Tracing App: A report to NHSX. The Conversation. 2020. URL: https://cdn.theconversation.com/static_files/files/1009/Report_-_Effective_App_Configurations.pdf?1587531217 [accessed 2021-01-25]
8. Walrave M, Waeterloos C, Ponnet K. Adoption of a Contact Tracing App for Containing COVID-19: A Health Belief Model Approach. JMIR Public Health Surveill 2020 Sep 01;6(3):e20572 [FREE Full text] [doi: 10.2196/20572] [Medline: 32755882]
9. von Wyl V, Höglinger M, Sieber C, Kaufmann M, Moser A, Serra-Burriel M, et al. Drivers of Acceptance of COVID-19 Proximity Tracing Apps in Switzerland: Panel Survey Analysis. JMIR Public Health Surveill 2021 Jan 06;7(1):e25701 [FREE Full text] [doi: 10.2196/25701] [Medline: 33326411]
10. Kaspar K. Motivations for Social Distancing and App Use as Complementary Measures to Combat the COVID-19 Pandemic: Quantitative Survey Study. J Med Internet Res 2020 Aug 27;22(8):e21613 [FREE Full text] [doi: 10.2196/21613] [Medline: 32759100]
11. Tomczyk S, Barth S, Schmidt S, Muehlan H. Utilizing Health Behavior Change and Technology Acceptance Models to Predict the Adoption of COVID-19 Contact Tracing Apps: Cross-sectional Survey Study. J Med Internet Res 2021 May 19;23(5):e25447 [FREE Full text] [doi: 10.2196/25447] [Medline: 33882016]
12. Trang S, Trenz M, Weiger WH, Tarafdar M, Cheung CM. One app to trace them all? Examining app specifications for mass acceptance of contact-tracing apps. European Journal of Information Systems 2020 Jul 27;29(4):415-428. [doi: 10.1080/0960085x.2020.1784046]
13. 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. The Lancet Global Health 2020 Apr;8(4):e488-e496 [FREE Full text] [doi: 10.1016/S2214-109X(20)30074-7] [Medline: 32119825]
14. Buitrago-Garcia D, Egli-Gany D, Counotte MJ, Hossmann S, Imeri H, Ipekci AM, et al. Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: A living systematic review and meta-analysis. PLoS Med 2020 Sep 22;17(9):e1003346 [FREE Full text] [doi: 10.1371/journal.pmed.1003346] [Medline: 32960881]
15. Braithwaite I, Callender T, Bullock M, Aldridge RW. Automated and partly automated contact tracing: a systematic review to inform the control of COVID-19. The Lancet Digital Health 2020 Nov;2(11):e607-e621. [doi: 10.1016/s2589-7500(20)30184-9]

16. Blom AG, Wenz A, Cornesse C, Rettig T, Fikel M, Friedel S, et al. Barriers to the Large-Scale Adoption of a COVID-19 Contact Tracing App in Germany: Survey Study. J Med Internet Res 2021 Mar 02;23(3):e23362 [FREE Full text] [doi: 10.2196/23362] [Medline: 33577466]
17. Dar AB, Lone AH, Zahoor S, Khan AA, Naaz R. Applicability of mobile contact tracing in fighting pandemic (COVID-19): Issues, challenges and solutions. Comput Sci Rev 2020 Nov;38:100307 [FREE Full text] [doi: 10.1016/j.cosrev.2020.100307] [Medline: 32989380]
18. Altmann S, Milsom L, Zillessen H, Blasone R, Gerdon F, Bach R, et al. Acceptability of App-Based Contact Tracing for COVID-19: Cross-Country Survey Study. JMIR Mhealth Uhealth 2020 Aug 28;8(8):e19857 [FREE Full text] [doi: 10.2196/19857] [Medline: 32759102]
19. Li H, Wu J, Gao Y, Shi Y. Examining individuals' adoption of healthcare wearable devices: An empirical study from privacy calculus perspective. Int J Med Inform 2016 Apr;88:8-17. [doi: 10.1016/j.ijmedinf.2015.12.010] [Medline: 26878757]
20. Princi E, Krämer NC. Out of Control - Privacy Calculus and the Effect of Perceived Control and Moral Considerations on the Usage of IoT Healthcare Devices. Front Psychol 2020 Nov 11;11:582054 [FREE Full text] [doi: 10.3389/fpsyg.2020.582054] [Medline: 33262731]
21. Seto E, Challa P, Ware P. Adoption of COVID-19 Contact Tracing Apps: A Balance Between Privacy and Effectiveness. J Med Internet Res 2021 Mar 04;23(3):e25726 [FREE Full text] [doi: 10.2196/25726] [Medline: 33617459]
22. Corona-Warn-App F.A.Q. Bundesregierung. 2020. URL: https://tinyurl.com/aa8u4s9y [accessed 2021-01-29]
23. Frequently Asked Questions about the Corona-Warn-App. Corona Warn-App Open Source Project. 2021. URL: https://www.coronawarn.app/en/faq/ [accessed 2021-01-29]
24. Federal Office of Public Health. Factsheet The SwissCovid app helps keep the coronavirus in check. Newsd. 2020. URL: https://www.newsd.admin.ch/newsd/message/attachments/61860.pdf [accessed 2021-01-29]
25. Anzahl aller bisher heruntergeladenen SwissCovid-Apps. Bundesamt für Statistik (BFS) - Experimentelle Statistiken. 2021. URL: https://www.experimental.bfs.admin.ch/expstat/de/home/innovative-methoden/swisscovid-app-monitoring.assetdetail.13367491.html [accessed 2021-03-29]
26. Kennzahlen zur Corona-Warn-App. Robert Koch Institut. 2021. URL: https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/WarnApp/Archiv_Kennzahlen/Kennzahlen_26032021.pdf?__blob=publicationFile [accessed 2021-03-29]
27. Tenzer F. Prognose zur Anzahl der Smartphone-Nutzer in der Schweiz von 2017 bis 2023. Statista. 2019. URL: https://de.statista.com/statistik/daten/studie/502662/umfrage/prognose-zur-anzahl-der-smartphonenutzer-in-der-schweiz/ [accessed 2021-03-29]
28. Tenzer F. Prognose zur Anzahl der Smartphone-Nutzer in Deutschland von 2017 bis 2023. Statista. 2019. URL: https://de.statista.com/statistik/daten/studie/500579/umfrage/prognose-zur-anzahl-der-smartphonenutzer-in-deutschland/ [accessed 2021-03-29]
29. Riemer K, Ciriello R, Peter S, Schlagwein D. Digital contact-tracing adoption in the COVID-19 pandemic: IT governance for collective action at the societal level. European Journal of Information Systems 2020 Sep 23;29(6):731-745. [doi: 10.1080/0960085X.2020.1819898]
30. Leppin A, Aro AR. Risk perceptions related to SARS and avian influenza: theoretical foundations of current empirical research. Int J Behav Med 2009 Feb 12;16(1):7-29 [FREE Full text] [doi: 10.1007/s12529-008-9002-8] [Medline: 19214752]
31. Floyd DL, Prentice-dunn S, Rogers RW. A Meta-Analysis of Research on Protection Motivation Theory. J Appl Social Pyschol 2000 Feb;30(2):407-429. [doi: 10.1111/j.1559-1816.2000.tb02323.x]
32. Rogers RW. A Protection Motivation Theory of Fear Appeals and Attitude Change. J Psychol 1975 Sep 02;91(1):93-114. [doi: 10.1080/00223980.1975.9915803] [Medline: 28136248]
33. Carpenter CJ. A meta-analysis of the effectiveness of health belief model variables in predicting behavior. Health Commun 2010 Dec 30;25(8):661-669. [doi: 10.1080/10410236.2010.521906] [Medline: 21153982]
34. Rosenstock IM. Why People Use Health Services. The Milbank Memorial Fund Quarterly 1966 Jul;44(3):94. [doi: 10.2307/3348967]
35. Bish A, Michie S. Demographic and attitudinal determinants of protective behaviours during a pandemic: A review. British Journal of Health Psychology 2010;15(4):797-824. [doi: 10.1348/135910710x485826]
36. Harper CA, Satchell LP, Fido D, Latzman RD. Functional Fear Predicts Public Health Compliance in the COVID-19 Pandemic. Int J Ment Health Addict 2020 Apr 27:1-14 [FREE Full text] [doi: 10.1007/s11469-020-00281-5] [Medline: 32346359]
37. Wise T, Zbozinek TD, Michelini G, Hagan CC, Mobbs D. Changes in risk perception and self-reported protective behaviour during the first week of the COVID-19 pandemic in the United States. R Soc Open Sci 2020 Sep 16;7(9):200742 [FREE Full text] [doi: 10.1098/rsos.200742] [Medline: 33047037]
38. Witte K. Putting the fear back into fear appeals: The extended parallel process model. Communication Monographs 1992 Dec;59(4):329-349. [doi: 10.1080/03637759209376276]
39. Tannenbaum MB, Hepler J, Zimmerman RS, Saul L, Jacobs S, Wilson K, et al. Appealing to fear: A meta-analysis of fear appeal effectiveness and theories. Psychol Bull 2015 Nov;141(6):1178-1204 [FREE Full text] [doi: 10.1037/a0039729] [Medline: 26501228]

40. Witte K, Allen M. A meta-analysis of fear appeals: implications for effective public health campaigns. Health Educ Behav 2000 Oct;27(5):591-615. [doi: 10.1177/109019810002700506] [Medline: 11009129]
41. Prati G, Pietrantoni L, Zani B. A social-cognitive model of pandemic influenza H1N1 risk perception and recommended behaviors in Italy. Risk Anal 2011 Apr;31(4):645-656. [doi: 10.1111/j.1539-6924.2010.01529.x] [Medline: 21077927]
42. 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;14:169-111 [FREE Full text] [doi: 10.1186/1471-2334-14-169] [Medline: 24674239]
43. Krösman C, Kriegeskotte N. 28 Millionen wollen Corona-Warn-App dauerhaft nutzen. Bitkom. 2020. URL: https://www.bitkom.org/Presse/Presseinformation/28-Millionen-wollen-Corona-Warn-App-dauerhaft-nutzen [accessed 2021-01-25]
44. Bosshardt L, Bühler G, Craviolini J, Hermann M, Krähenbühl D. Tracing App des Bundes - Studienbericht zur Bevölkerungsbefragung. Newsd. 2020. URL: https://www.newsd.admin.ch/newsd/message/attachments/61463.pdf [accessed 2021-01-25]
45. Garrett P, White J, Lewandowsky S, Kashima Y, Perfors A, Little D, et al. The acceptability and uptake of smartphone tracking for COVID-19 in Australia. PLoS One 2021;16(1):e0244827 [FREE Full text] [doi: 10.1371/journal.pone.0244827] [Medline: 33481841]
46. Corona Warn-App - KW 48/2020. KANTAR. 2020. URL: http://docs.dpaq.de/17221-kantar_umfrage_corona-app.pdf [accessed 2021-01-25]
47. O'Callaghan ME, Buckley J, Fitzgerald B, Johnson K, Laffey J, McNicholas B, et al. A national survey of attitudes to COVID-19 digital contact tracing in the Republic of Ireland. Ir J Med Sci 2021 Aug 16;190(3):863-887 [FREE Full text] [doi: 10.1007/s11845-020-02389-y] [Medline: 33063226]
48. Guillon M, Kergall P. Attitudes and opinions on quarantine and support for a contact-tracing application in France during the COVID-19 outbreak. Public Health 2020 Nov;188:21-31 [FREE Full text] [doi: 10.1016/j.puhe.2020.08.026] [Medline: 33059232]
49. Jonker M, de Bekker-Grob E, Veldwijk J, Goossens L, Bour S, Rutten-Van Mölken M. COVID-19 Contact Tracing Apps: Predicted Uptake in the Netherlands Based on a Discrete Choice Experiment. JMIR Mhealth Uhealth 2020 Oct 09;8(10):e20741 [FREE Full text] [doi: 10.2196/20741] [Medline: 32795998]
50. Mouter N, Collewet M, de Wit GA, Rotteveel A, Lambooij MS, Kessels R. Societal Effects Are a Major Factor for the Uptake of the Coronavirus Disease 2019 (COVID-19) Digital Contact Tracing App in The Netherlands. Value Health 2021 May;24(5):658-667 [FREE Full text] [doi: 10.1016/j.jval.2021.01.001] [Medline: 33933234]
51. Van Vugt M, Van Lange P. The Altruism Puzzle: Psychological Adaptations for Prosocial Behavior. In: Schaller M, Simpson JA, Kenrick DT, editors. Evolution and Social Psychology. London, UK: Psychology Press; 2006:237-261.
52. O’Connor AM, Evans AD. Dishonesty during a pandemic: The concealment of COVID-19 information. J Health Psychol 2020 Aug 17:135910532095160. [doi: 10.1177/1359105320951603]
53. Pfattheicher S, Nockur L, Böhm R, Sassenrath C, Petersen MB. The Emotional Path to Action: Empathy Promotes Physical Distancing and Wearing of Face Masks During the COVID-19 Pandemic. Psychol Sci 2020 Nov 29;31(11):1363-1373. [doi: 10.1177/0956797620964422] [Medline: 32993455]
54. Pfattheicher S, Petersen MB, Böhm R. Information about herd immunity through vaccination and empathy promote COVID-19 vaccination intentions. PsyArXiv. URL: https://psyarxiv.com/wzu6k/ [accessed 2021-03-29]
55. Webb TL, Sheeran P. Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. Psychol Bull 2006 Mar;132(2):249-268. [doi: 10.1037/0033-2909.132.2.249] [Medline: 16536643]
56. Oldeweme A, Märtins J, Westmattelmann D, Schewe G. The Role of Transparency, Trust, and Social Influence on Uncertainty Reduction in Times of Pandemics: Empirical Study on the Adoption of COVID-19 Tracing Apps. J Med Internet Res 2021 Feb 08;23(2):e25893 [FREE Full text] [doi: 10.2196/25893] [Medline: 33465036]
57. Randall DM, Fernandes MF. The social desirability response bias in ethics research. J Bus Ethics 1991 Nov;10(11):805-817. [doi: 10.1007/BF00383696]
58. Imhoff R, Lamberty P. A Bioweapon or a Hoax? The Link Between Distinct Conspiracy Beliefs About the Coronavirus Disease (COVID-19) Outbreak and Pandemic Behavior. Social Psychological and Personality Science 2020 Jul 06;11(8):1110-1118. [doi: 10.1177/1948550620934692]
59. Min C, Shen F, Yu W, Chu Y. The relationship between government trust and preventive behaviors during the COVID-19 pandemic in China: Exploring the roles of knowledge and negative emotion. Prev Med 2020 Dec;141:106288 [FREE Full text] [doi: 10.1016/j.ypmed.2020.106288] [Medline: 33091414]
60. 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]
61. Dryhurst S, Schneider CR, Kerr J, Freeman ALJ, Recchia G, van der Bles AM, et al. Risk perceptions of COVID-19 around the world. Journal of Risk Research 2020 May 05;23(7-8):994-1006. [doi: 10.1080/13669877.2020.1758193]
62. Huber M, Langen H. Timing matters: the impact of response measures on COVID-19-related hospitalization and death rates in Germany and Switzerland. Swiss J Econ Stat 2020 Aug 26;156(1):10 [FREE Full text] [doi: 10.1186/s41937-020-00054-w] [Medline: 32864361]

63. Saw YE, Tan EY, Liu JS, Liu JC. Predicting Public Uptake of Digital Contact Tracing During the COVID-19 Pandemic: Results From a Nationwide Survey in Singapore. J Med Internet Res 2021 Feb 03;23(2):e24730 [FREE Full text] [doi: 10.2196/24730] [Medline: 33465034]
64. Brewer NT, Weinstein ND, Cuite CL, Herrington JE. Risk perceptions and their relation to risk behavior. Ann Behav Med 2004 Apr;27(2):125-130. [doi: 10.1207/s15324796abm2702_7] [Medline: 15026296]
65. Qualtrics. URL: https://www.qualtrics.com/ [accessed 2021-07-23]
66. Ritchie H, Ortiz-Ospina E, Beltekian D, Mathieu E, Hasell J, Macdonald B. Policy Responses to the Coronavirus Pandemic. Our World in Data. 2021. URL: https://ourworldindata.org/policy-responses-covid [accessed 2021-05-27]
67. Ritchie H, Ortiz-Ospina E, Beltekian D, Mathieu E, Hasell J, Macdonald B. Coronavirus (COVID-19) Cases. Our World in Data. 2021. URL: https://ourworldindata.org/covid-cases [accessed 2021-05-27]
68. Ritchie H, Ortiz-Ospina E, Beltekian D, Mathieu E, Hasell J, Macdonald B. Coronavirus (COVID-19) Deaths. Our World in Data. 2021. URL: https://ourworldindata.org/covid-deaths [accessed 2021-05-27]
69. Nationalratswahlen 2019 - Wähleranteil der wichtigsten Parteien (in %). Bundesamt für Statistik. 2019. URL: https://www.bfs.admin.ch/bfs/de/home/statistiken/politik.html [accessed 2021-01-25]
70. Bildungsstand der ständigen Wohnbevölkerung nach Alter und Geschlecht. Bundesamt für Statistik. 2020. URL: https://www.bfs.admin.ch/bfs/de/home/statistiken/bildung-wissenschaft/bildungsstand.assetdetail.12527184.html [accessed 2021-01-25]
71. Ständige Wohnbevölkerung nach Staatsangehörigkeitskategorie und Geschlecht, 1950-2019. Bundesamt für Statistik. 2020. URL: https://www.bfs.admin.ch/bfs/de/home/statistiken/bevoelkerung/stand-entwicklung/alter-zivilstand-staatsangehoerigkeit.assetdetail.13707235.html [accessed 2021-01-25]
72. Deutsche Bevölkerung 2011 bis 2018 Grund- und Verhältniszahlen zum allgemeinbildenden Schulabschluss. Statistisches Bundesamt. 2019. URL: https://tinyurl.com/3knj3twv [accessed 2021-01-25]
73. Bevölkerung (Zensus): Deutschland, Stichtag, Nationalität, Geschlecht. Statistisches Bundesamt. 2020. URL: https://tinyurl.com/yjudfd7f [accessed 2021-01-25]
74. Gültige Zweitstimmen (Repräsentative Bundestagswahlstatistik): Deutschland, Stichtag, Parteien, Geschlecht. Statistisches Bundesamt. 2020. URL: https://tinyurl.com/rr2efdxn [accessed 2021-01-25]
75. R Foundation for Statistical Computing. The R Project for Statistical Computing. R Project. 2020. URL: https://www.R-project.org/ [accessed 2020-09-22]
76. Ranganathan P, Pramesh C, Aggarwal R. Common pitfalls in statistical analysis: Logistic regression. Perspect Clin Res 2017;8(3):148-151 [FREE Full text] [doi: 10.4103/picr.PICR_87_17] [Medline: 28828311]
77. Peng CJ, So TH. Logistic Regression Analysis and Reporting: A Primer. Understanding Statistics 2002 Feb 02;1(1):31-70. [doi: 10.1207/s15328031us0101_04]
78. Ottenbacher KJ, Ottenbacher HR, Tooth L, Ostir GV. A review of two journals found that articles using multivariable logistic regression frequently did not report commonly recommended assumptions. J Clin Epidemiol 2004 Nov;57(11):1147-1152. [doi: 10.1016/j.jclinepi.2003.05.003] [Medline: 15567630]
79. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. Journal of Clinical Epidemiology 1996 Dec;49(12):1373-1379. [doi: 10.1016/S0895-4356(96)00236-3]
80. Hosmer D, Lemeshow S. Applied Logistic Regression. Hoboken, NJ: Wiley; 2000.
81. Thomas R, Michaleff ZA, Greenwood H, Abukmail E, Glasziou P. Concerns and Misconceptions About the Australian Government's COVIDSafe App: Cross-Sectional Survey Study. JMIR Public Health Surveill 2020 Nov 04;6(4):e23081 [FREE Full text] [doi: 10.2196/23081] [Medline: 33048826]
82. Zimmermann BM, Fiske A, Prainsack B, Hangel N, McLennan S, Buyx A. Early Perceptions of COVID-19 Contact Tracing Apps in German-Speaking Countries: Comparative Mixed Methods Study. J Med Internet Res 2021 Feb 08;23(2):e25525 [FREE Full text] [doi: 10.2196/25525] [Medline: 33503000]
83. Theobald T. Werbespot Corona-Warn-App. Horizont. 2020 Jun 17. URL: https://www.horizont.net/marketing/nachrichten/corona-warn-app-das-sind-alle-motive-der-kampagne-183691 [accessed 2020-10-23]
84. Fergus TA, Bardeen JR. Anxiety sensitivity and intolerance of uncertainty: Evidence of incremental specificity in relation to health anxiety. Personality and Individual Differences 2013 Oct;55(6):640-644. [doi: 10.1016/j.paid.2013.05.016]
85. Kawakami N, Sasaki N, Kuroda R, Tsuno K, Imamura K. The Effects of Downloading a Government-Issued COVID-19 Contact Tracing App on Psychological Distress During the Pandemic Among Employed Adults: Prospective Study. JMIR Ment Health 2021 Jan 12;8(1):e23699 [FREE Full text] [doi: 10.2196/23699] [Medline: 33347424]
86. Vehovar V, Toepoel V, Steinmetz S. Non-probability sampling. In: Wolf C, Joye D, Smith TW, Fu YC, editors. The Sage Handbook of Survey Methodology. Thousand Oaks, CA: Sage Publications; 2016:329-345.
87. Shadish WR, Cook TD, Campbell DT. Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Boston, MA: Cengage Learning; 2002.
88. Furr M. Scale Construction and Psychometrics for Social and Personality Psychology. Thousand Oaks, CA: SAGE Publications; 2011.

89. Nichols AL, Webster GD. The single-item need to belong scale. Personality and Individual Differences 2013 Jul;55(2):189-192. [doi: 10.1016/j.paid.2013.02.018]
90. Hays J. Epidemics and Pandemics: Their Impacts on Human History. Santa Barbara, CA: ABC-Clio; 2005.
91. Cooper BS, Pitman RJ, Edmunds WJ, Gay NJ. Delaying the international spread of pandemic influenza. PLoS Med 2006 Jun 2;3(6):e212 [FREE Full text] [doi: 10.1371/journal.pmed.0030212] [Medline: 16640458]
92. Platto S, Zhou J, Wang Y, Wang H, Carafoli E. Biodiversity loss and COVID-19 pandemic: The role of bats in the origin and the spreading of the disease. Biochem Biophys Res Commun 2021 Jan 29;538:2-13 [FREE Full text] [doi: 10.1016/j.bbrc.2020.10.028] [Medline: 33092787]

Abbreviations

AUC: area under the receiver operating characteristic curve
CTA: contact tracing app
OR: odds ratio

Edited by C Basch; submitted 31.03.21; peer-reviewed by M Walrave, I Mircheva; comments to author 15.05.21; accepted 04.06.21; published 10.08.21.

Please cite as:
große Deters F, Meier T, Milek A, Horn AB
Self-Focused and Other-Focused Health Concerns as Predictors of the Uptake of Corona Contact Tracing Apps: Empirical Study
J Med Internet Res 2021;23(8):e29268
URL: https://www.jmir.org/2021/8/e29268
doi:10.2196/29268
PMID:34227995

©Fenne große Deters, Tabea Meier, Anne Milek, Andrea B Horn. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Leveraging Transfer Learning to Analyze Opinions, Attitudes, and Behavioral Intentions Toward COVID-19 Vaccines: Social Media Content and Temporal Analysis

Siru Liu[1], PhD; Jili Li[2], BS; Jialin Liu[3], MD

[1]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, United States

[2]West China Medical School, Sichuan University, Chengdu, China

[3]Department of Medical Informatics, West China Hospital, Sichuan University, Chengdu, China

Corresponding Author:
Jialin Liu, MD
Department of Medical Informatics
West China Hospital
Sichuan University
No 37 Wainan Guoxuexiang Street
Chengdu, 610041
China
Phone: 86 28 85422306
Email: dljl8@163.com

Abstract

Background: The COVID-19 vaccine is considered to be the most promising approach to alleviate the pandemic. However, in recent surveys, acceptance of the COVID-19 vaccine has been low. To design more effective outreach interventions, there is an urgent need to understand public perceptions of COVID-19 vaccines.

Objective: Our objective was to analyze the potential of leveraging transfer learning to detect tweets containing opinions, attitudes, and behavioral intentions toward COVID-19 vaccines, and to explore temporal trends as well as automatically extract topics across a large number of tweets.

Methods: We developed machine learning and transfer learning models to classify tweets, followed by temporal analysis and topic modeling on a dataset of COVID-19 vaccine–related tweets posted from November 1, 2020 to January 31, 2021. We used the F1 values as the primary outcome to compare the performance of machine learning and transfer learning models. The statistical values and *P* values from the Augmented Dickey-Fuller test were used to assess whether users' perceptions changed over time. The main topics in tweets were extracted by latent Dirichlet allocation analysis.

Results: We collected 2,678,372 tweets related to COVID-19 vaccines from 841,978 unique users and annotated 5000 tweets. The F1 values of transfer learning models were 0.792 (95% CI 0.789-0.795), 0.578 (95% CI 0.572-0.584), and 0.614 (95% CI 0.606-0.622) for these three tasks, which significantly outperformed the machine learning models (logistic regression, random forest, and support vector machine). The prevalence of tweets containing attitudes and behavioral intentions varied significantly over time. Specifically, tweets containing positive behavioral intentions increased significantly in December 2020. In addition, we selected tweets in the following categories: positive attitudes, negative attitudes, positive behavioral intentions, and negative behavioral intentions. We then identified 10 main topics and relevant terms for each category.

Conclusions: Overall, we provided a method to automatically analyze the public understanding of COVID-19 vaccines from real-time data in social media, which can be used to tailor educational programs and other interventions to effectively promote the public acceptance of COVID-19 vaccines.

(J Med Internet Res 2021;23(8):e30251) doi:10.2196/30251

KEYWORDS

vaccine; COVID-19; leveraging transfer learning; pandemic; infodemiology; infoveillance; public health; social media; content analysis; machine learning; online health

Introduction

The outbreak of COVID-19 has affected 219 countries and territories with 102,083,344 confirmed cases causing 2,209,195 deaths as of January 31, 2021, as reported by the World Health Organization (WHO) [1]. As a significant global health threat, long-term control of COVID-19 relies on the development and acceptance of a preventive vaccine [2-4]. Fortunately, in November 2020, Pfizer-BioNTech and Moderna reported more than 95% efficacy of their vaccines [5], which were subsequently authorized by the US Food and Drug Administration (FDA) for emergency use. Since the preventive vaccine has been successfully developed, the current barrier is obtaining a sufficient proportion of the population to accept vaccines to slow the spread of the outbreak [6]. However, according to a recent survey, only 51% of 10,093 adults in the United States indicated that they would be willing to receive the COVID-19 vaccine when it becomes available [7], which would not achieve the recommended threshold of 70% to reach herd immunity [8].

Vaccine hesitancy, defined as "a behavior with delay in acceptance or refusal of vaccines despite available services," was identified by the WHO as a global threat in 2019 [9]. The SAGA Working Group developed the Vaccine Hesitancy Determinant Matrix, including contextual influences (ie, related to historic, sociocultural, environmental, institutional, economic, or political factors), individual and group influences (ie, factors related to personal perception or social environment), and vaccine/vaccination-specific issues [10]. Unlike other common vaccines, the COVID-19 vaccines are associated with many factors that might amplify vaccine hesitancy [11,12]. Previous studies have reported widespread public concern about the rapid speed of vaccine development, novelty of the development technology (mRNA), unknown long-term side effects, and politicization of vaccines [13,14]. Furthermore, the social environment is polarized, with distrust of science among some groups and a plethora of conspiracies and misinformation about vaccines spreading across social media platforms [15,16]. For these reasons, it might be more difficult to achieve the coverage goal for COVID-19 vaccines. Therefore, it is urgent to efficiently collect information on public perceptions to tailor education materials for public and clinical guidance, which will enable primary care physicians to promote COVID-19 vaccines.

With the increased growth of internet-based applications, more people have begun sharing their opinions on social media platforms. In particular, during the current COVID-19 pandemic, people may increase their use of social media due to social distancing [11]. Social media is awash with virus conspiracies and misinformation [15]. Various social media platforms (eg, Facebook, Instagram, Reddit) are currently providing health information to researchers; among them, the Twitter platform has a more prominent role in gathering public perceptions on health care [17]. Twitter has become a good data source to collect real-time perceptions from a large-scale population for public health research. Over the past decades, researchers have used social media analytics tools to monitor public sentiment and communication patterns in a global pandemic crisis (eg, Ebola and Zika outbreaks) [18-20]. Mavragani [21] performed a time-series analysis on Google trends data and found a significant correlation between search interests with reported COVID-19 cases. Li et al [22] developed a taxonomy of Weibo posts on COVID-19 topics, and Liao et al [23] analyzed Weibo posts to identify public engagement and government responsiveness. Fadda et al [24] performed a content analysis to examine the extent of vaccine conspiracy theories reflected in tweets. Our study focused on the behavioral intentions related to COVID-19 vaccines, which is different from previous studies that performed a general analysis of COVID-19 tweets or vaccine conspiracy theories. Findings of this study could directly help researchers and policymakers to develop more targeted implementation strategies to improve acceptance rates of COVID-19 vaccines.

Machine learning and deep learning techniques have been used as efficient methods to detect public perceptions on social media platforms. In health care, researchers have developed deep learning models to perform longitudinal and geographic analyses to understand human papillomavirus (HPV) vaccine discussions [25]. These models also achieved good performance in predicting diagnosis or identifying patients in a high-risk group [26-28]. Transfer learning, as an emerging deep learning technique, has been applied to classify computed tomography images and notes. In transfer learning, a pretrained model is first used, which is then fine-tuned based on the specific datasets and tasks. Because the pretrained model already contains large-scale domain knowledge, the classification performance can achieve high values even with fine-tuning on relatively small datasets [29]. In this study, we applied Google's bidirectional encoder representations from transformers (BERT) model as the pretrained model, which has achieved new state-of-the-art results in the natural language processing domain [29].

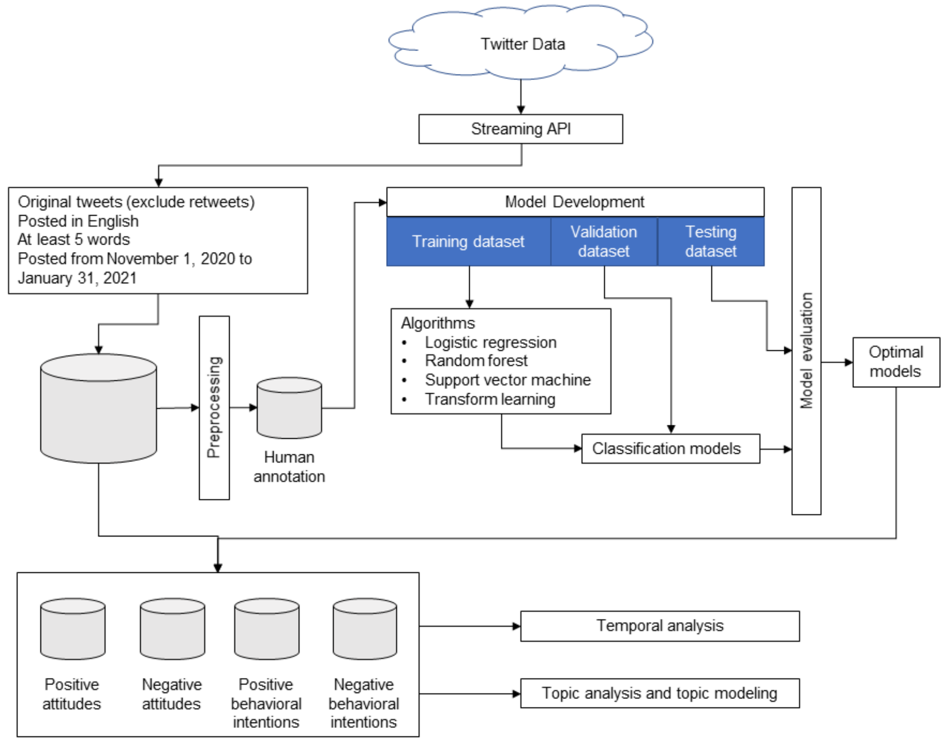
Although previous studies have explored additional knowledge in the context of other vaccines using machine learning and deep learning methods, several questions related to COVID-19 vaccines remain unknown: What is the prevalence of user opinions on a social media platform? How many tweets express positive/negative attitudes and behavioral intentions to take vaccines? Which topics are mostly associated with these contents? To answer these questions, we developed machine learning models (logistic regression, random forest, support vector machine) and transfer learning models to detect the content expressing user opinions, attitudes, and behavioral intentions toward COVID-19 vaccines. We then performed a temporal analysis to explore trends over time and developed probabilistic topic models to obtain the most important and valuable topics. We believe that this study will be of great benefit to the timely rollout of COVID-19 vaccines by extracting the latest public opinions, attitudes, and behavioral intentions that can help tailor promotion programs to fit different populations.

Methods

Study Overview

We collected tweets related to COVID-19 vaccines posted from November 1, 2020 to January 31, 2021, and annotated 5000 tweets as the gold standard. We developed machine learning and transfer learning models to classify tweets for three tasks: (1) opinions (yes, no); (2) attitudes (positive, negative, neutral); and (3) behavioral intentions (positive, negative, unknown). The above tasks all focused on COVID-19 vaccines. We then applied the models to predict unlabeled tweets and performed a temporal analysis to capture trends in the unlabeled tweets. In addition, we performed a topic analysis using word clouds and a latent Dirichlet allocation (LDA) model to further understand the content of tweets in the following categories: positive attitudes, negative attitudes, positive behavioral intentions, and negative behavioral intentions. The overall framework is shown in Figure 1.

Figure 1. Overall study framework. API: application programming interface.


Twitter Data
Streaming API
Original tweets (exclude retweets)
Posted in English
At least 5 words
Posted from November 1, 2020 to January 31, 2021
Model Development
Training dataset
Validation dataset
Testing dataset
Algorithms
• Logistic regression
• Random forest
• Support vector machine
• Transform learning
Model evaluation
Optimal models
Preprocessing
Human annotation
Classification models
Positive attitudes
Negative attitudes
Positive behavioral intentions
Negative behavioral intentions
Temporal analysis
Topic analysis and topic modeling

Data Collection

We used a combination of keywords and hashtags related to COVID-19 vaccines to collect tweets in English published from November 1, 2020 to January 31, 2021. We intentionally chose November, following the announcement of the first effective vaccine on November 9, 2020, to determine if the announcement of successful vaccine trial results might influence the perceptions of vaccines or vaccination. The search strategy employed the following search terms: "(#covid OR covid OR #covid19 OR covid19) AND (#vaccine OR vaccine OR #vacine OR vacine OR vaccinate OR immunization OR immune OR vax) since:2020-11-01 until:2021-01-31 lang:en." We used snscrape and tweepy in Python 3 to collect data and to exclude retweets. To clean up the original tweets, we removed nonalphanumeric characters and converted the text to lowercase. We randomly selected 5000 tweets from November 1, 2020 to November 22, 2020, annotated by two independent reviewers (SL and JL) in batches of 200. Any annotation disagreements were discussed and adjudicated by the supervising investigators. For each tweet, we first labeled whether it included a user opinion toward the COVID-19 vaccines (yes or no). We considered a tweet to include an opinion about the COVID-19 vaccines if it met both of the following conditions: (1) targeted at the COVID-19 vaccines and (2) generated by a user. For the tweets that expressed user opinions toward the COVID-19 vaccines, we labeled the attitude (positive, negative, or neutral) and the behavioral intention (positive, negative, or unknown) toward COVID-19 vaccines. The attitude category used the traditional emotional polarity. The analysis of attitude was performed on the aspect level. If both positive and negative attitudes toward COVID-19 vaccines were present in the same tweet, we labeled it in the unknown category. The coding rules were iteratively developed by our group in which an independent review was performed, disagreements were discussed, and coding rules were revised. This process continued until the interrater agreement reached ≥0.80. The annotated corpus was used as a gold standard to train and evaluate the machine learning and transfer learning models.

Model Development and Evaluation

For data preprocessing, we used the tweet-preprocessor package in Python 3 to remove URLs, hashtags, mentions, reserved words (eg, RT, FAV), emojis, smileys, and numbers in each tweet. We split the annotated dataset into three parts: training (60%), validation (20%), and testing (20%). The training and validation datasets were used to train models and select optimal hyperparameters through 5-fold cross-validation. We applied transfer learning using text frequency-inverse document frequency to compare traditional machine learning algorithms (logistic regression, random forest, and support vector machine) to transfer learning models. The machine learning models were developed using the scikit-learn package in Python 3.

For transfer learning, we used the BERT-base-cased as the pretrained language model and the “BERT for sequence classification” model as the pretrained classification model. Because the BERT model requires each sentence to be the same length, we padded each tweet with 64 tokens, as most tweets have lengths in this range. We then fine-tuned this model on the training and validation datasets using the Adam algorithm with weight decay (AdamW) as an optimizer. We performed three text classification tasks. We first developed a binary classifier to determine whether the tweets state an opinion related to the COVID-19 vaccines. We then developed two multiclass classifiers to categorize attitudes and behavioral intentions, respectively. The BERT models were generated using the huggingface package in Python 3. The models were developed with the Google Colab platform using a high-RAM GPU.

We evaluated the models on the testing dataset and report outcomes with 1000 rounds of bootstrapping. The primary outcome was the macro-F1 value and the secondary outcomes were recall, precision, and accuracy. We performed the Nemenyi test to compare the F1 values of traditional machine learning models and transfer learning models [30]. The model with the highest F1 value was considered the optimal model.

Temporal Analysis

We applied the optimal models to predict the unlabeled data for 3 months starting from November 1, 2020. For the task of extracting opinions, we calculated the proportion of tweets classified as containing opinions to the total number of tweets posted each day about the COVID-19 vaccines. For the tasks of classifying attitudes and behavioral intentions toward the COVID-19 vaccines, we calculated the percentage of tweets predicted to exhibit a particular attitude or behavioral intention to all tweets indicating attitudes or behavioral intentions, respectively. To assess the statistical significance of variability over time, we performed the Augmented Dickey-Fuller (ADF) test [31] with a significance threshold of *P*<.05. The ADF test is a unit root test, which is commonly used to determine the stationarity of a time-series sample.

Topic Analysis and Topic Modeling

To understand the content of tweets in each category, we used word clouds to illustrate the frequency of words appearing in the content. The more frequently used words have larger sizes, indicating more importance in the category [32]. Furthermore, we performed the LDA analysis to extract the main topics of discussion. LDA is a widely used unsupervised method that automatically clusters text based on content and identifies keywords in each topic through a probabilistic model [33,34]. We performed 5-fold cross-validation to tune hyperparameters in the LDA model (number of components and learning rate). After obtaining the results of the LDA models, we visualized extracted topics using the pyLDAvis library [35] in Python 3, which is an interactive visualization tool for displaying the distribution of topics and the top 30 most relevant terms with their weights in each topic.

Results

Performance of Classification Models

We annotated 5000 tweets from 4796 unique users with an average interrater reliability (κ) of 0.76. The prediction performances of models on the testing dataset using four different algorithms for three tasks are presented in Table 1. The transfer learning model significantly outperformed the machine learning models in identifying tweets that included opinions, attitudes, and behavioral intentions, achieving the highest F1 values of 0.792, 0.578, and 0.614, respectively.

Table 1. Metrics of transfer learning models and machine learning models in classifying tweets related to COVID-19 vaccines.

Task	Recall, mean (95% CI)	Precision, mean (95% CI)	F1, mean (95% CI)	Accuracy, mean (95% CI)
Opinions				
BERT[a]	0.762 (0.759-0.766)	0.862 (0.858-0.866)	0.792[b] (0.789-0.795)	0.854 (0.852-0.856)
Logistic regression	0.774 (0.770-0.779)	0.757 (0.753-0.762)	0.764 (0.761-0.767)	0.807 (0.805-0.810)
Random forest	0.754 (0.750-0.758)	0.732 (0.728-0.735)	0.740 (0.737-0.743)	0.783 (0.781-0.786)
Support vector machine	0.767 (0.764-0.771)	0.752 (0.748-0.755)	0.758 (0.755-0.761)	0.803 (0.801-0.806)
Attitudes				
BERT	0.529 (0.521-0.536)	0.698 (0.686-0.710)	0.578[b] (0.572-0.584)	0.873 (0.871-0.875)
Logistic regression	0.475 (0.468-0.482)	0.530 (0.520-0.541)	0.495 (0.490-0.500)	0.859 (0.856-0.861)
Random forest	0.518 (0.511-0.526)	0.558 (0.545-0.570)	0.508 (0.502-0.514)	0.830 (0.827-0.833)
Support vector machine	0.506 (0.498-0.514)	0.551 (0.541-0.562)	0.523 (0.517-0.530)	0.863 (0.860-0.865)
Behavioral intentions				
BERT	0.562 (0.549-0.575)	0.734 (0.716-0.752)	0.614[b] (0.606-0.622)	0.961 (0.960-0.962)
Logistic regression	0.472 (0.461-0.483)	0.725 (0.699-0.752)	0.527 (0.519-0.536)	0.951 (0.949-0.952)
Random forest	0.447 (0.437-0.457)	0.577 (0.543-0.611]	0.466 (0.457-0.476)	0.935 (0.934-0.937)
Support vector machine	0.469 (0.458-0.479)	0.710 (0.684-0.737)	0.523 (0.513-0.533)	0.950 (0.948-0.951)

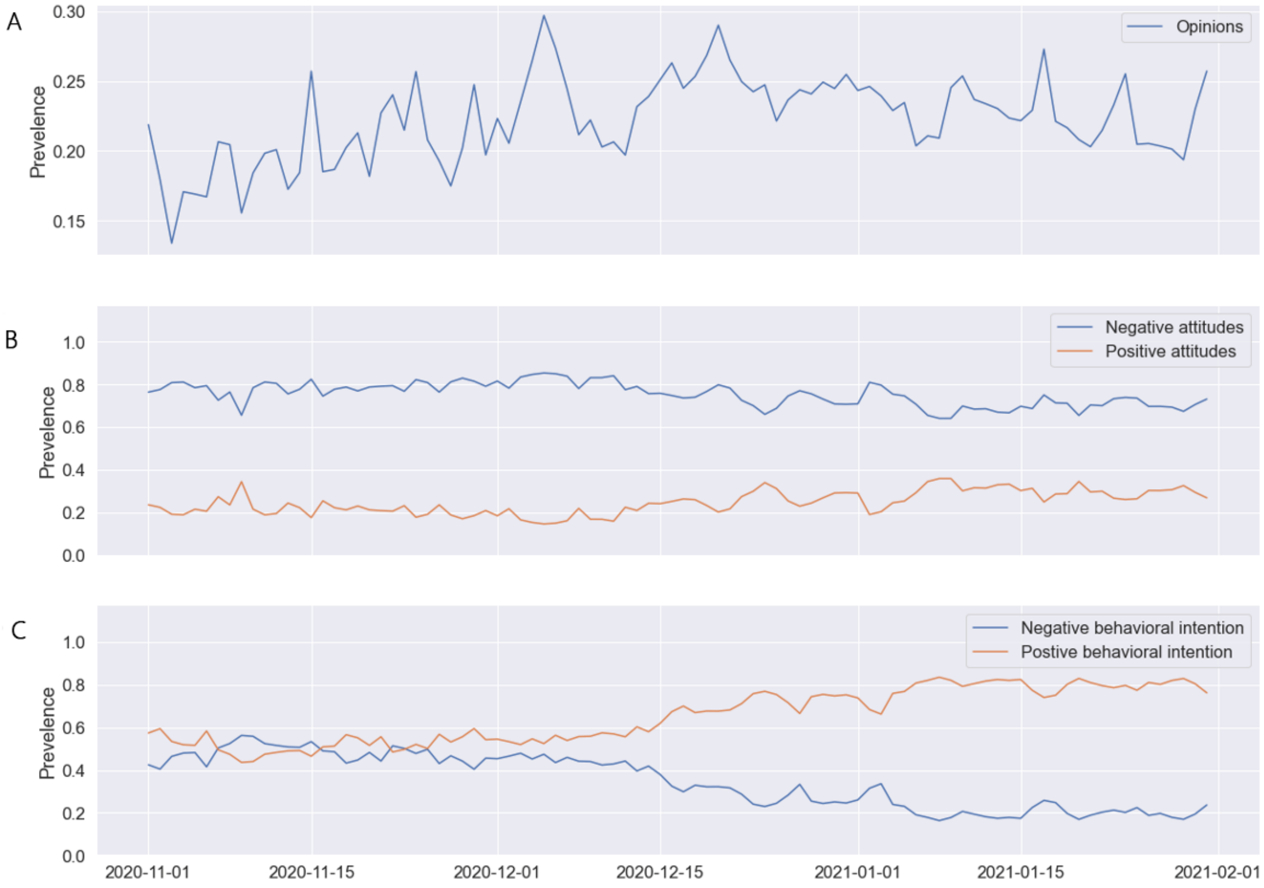
[a] BERT: Bidirectional encoder representations from transformers.

[b] *P*=.001 in the Nemenyi test.

Temporal Analysis

We collected 2,678,372 tweets related to COVID-19 vaccines posted by 841,978 unique users from November 1, 2020 to January 31, 2021. The daily prevalence distributions of opinions, attitudes, and behavioral intentions are shown in Figure 2. The daily prevalence of tweets expressing users' opinions was 0.222 (95% CI 0.202-0.245). The ADF statistic was –4.341 (*P*<.001), indicating that the time-series data were stationary. This reflects that the prevalence of tweets expressing opinions did not change significantly over time. For tweets containing attitudes toward the COVID-19 vaccines, the rate of negative attitudes was 0.754 (95% CI 0.707-0.795), while the rate of positive attitudes was only 0.246 (95% CI 0.204-0.293). The daily prevalence of attitudes was nonstationary (ADF –1.137, *P*=.70), which indicated a significant change in users' attitudes toward vaccines over time. Among tweets related to behavioral intentions, the rate of tweets indicating that users will not get vaccinated was 0.342 (95% CI 0.229-0.461), whereas the rate of tweets indicating that users will get vaccinated was 0.652 (95% CI 0.539-0.771). The behavioral intention prevalence was also nonstationary (ADF –0.980, *P*=.76), indicating that it varied significantly over time. Notably, we observed a substantial increase in the prevalence of tweets expressing positive behavioral intention starting from mid-December 2020.

Figure 2. Distribution of the prevalence of the tweets containing opinions (A), attitudes (B), and behavioral intentions (C) about COVID-19 vaccines for each day from November 1, 2020 to January 31, 2021.

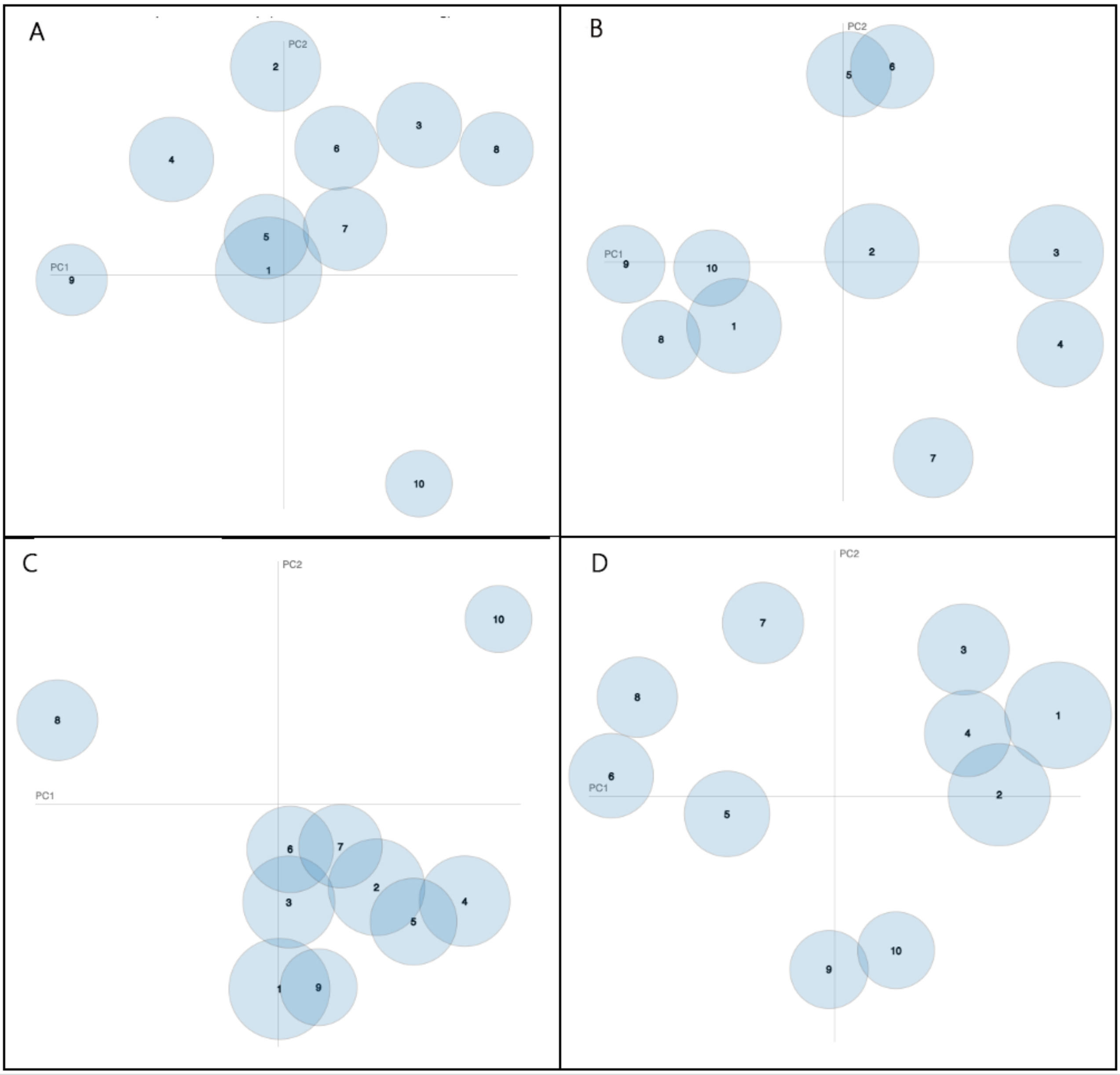


Topic Modeling and Analysis

Primary Domain Topics

After tuning hyperparameters of the LDA models, each model had 10 components (topics). Figure 3 presents intertopic distance maps generated by tweets containing positive/negative attitudes and positive/negative behavioral intentions. The size of bubbles represents the ratio of relevant tweets in that topic to the total number of tweets. In the following sections, we selected several domain topics for tweets in each category to describe the potential inferred themes based on identified relevant key terms. The overall top 15 keywords for each topic are listed in Textbox 1.

Figure 3. Intertopic distance maps for tweets that contained information in the following categories: negative attitudes (A), positive attitudes (B), negative behavioral intentions (C), and positive behavioral intentions (D).



Textbox 1. Top 15 key terms for each topic.

Negative attitudes

1. worry, prevent, covid, stop, need, spread, symptom, transmission, catch, people, reduce, infection, virus, eat, doesn
2. death, covid, case, people, rate, die, number, cause, population, test, trial, fear, report, survival, day
3. risk, covid, test, people, health, worker, trial, know, need, woman, work, child, pregnant, safe, age
4. effect, long, term, know, covid, bad, unknown, risk, affect, people, study, concern, damage, potential, impact
5. covid, year, make, anti, month, mask, rush, people, want, safe, need, just, know, sense, wear
6. covid, dose, use, virus, immune, antibody, body, immunity, trial, second, make, protein, cell, test, response
7. virus, new, covid, strain, effective, work, mutate, year, develop, mutation, research, cold, variant, different, make
8. covid, people, just, say, think, make, know, trust, want, cure, government, believe, thing, come
9. covid, die, people, life, chance, treatment, old, kill, effective, want, say, sick, save, safe, family
10. flu, covid, reaction, shot, drug, adverse, expect, people, shoot, allergic, just, high, bad, year, polio

Positive attitudes

1. covid, thank, work, great, today, day, make, worker, scientist, happy, mom, care, just, hard, nurse
2. covid, feel, effect, day, long, arm, just, little, work, fine, hour, term, good, excited, sore
3. safe, stay, end, covid, news, pandemic, effective, trial, good, amp, light, continue, home, hope, step
4. covid, hope, soon, look, forward, normal, life, hopefully, come, available, new, world, news, return, year
5. covid, good, year, just, time, wait, thing, hope, think, come, pray, love, wish, news, day
6. people, covid, want, need, know, die, risk, just, really, say, think, make, life, safe, fear
7. covid, dose, receive, today, grateful, second, family, feel, patient, able, thankful, protect, friend, happy, excited
8. flu, virus, covid, make, immune, fight, sure, body, new, immunity, just, strain, world, distribute, cause
9. mask, wear, covid, stop, social, spread, distancing, hand, catch, need, distance, people, virus, stay, help
10. covid, vaccinate, amp, case, symptom, prevent, ready, immunity, just, mean, virus, reduce, life, rate, infection

Negative behavioral intentions

1. covid, virus, stop, prevent, symptom, test, dose, immune, spread, mask, antibody, sick, just, catch, body
2. covid, flu, shot, shit, shoot, just, allow, work, win, scare, dead, year, virus, arm, sure
3. risk, covid, say, immune, make, high, virus, people, disease, just, healthy, sense, dangerous, case, good
4. want, covid, vaccinate, child, use, kill, kid, new, wait, way, cure, effective, doctor, just, people
5. covid, body, rate, vaccination, survival, choice, eat, mandatory, know, worry, life, fear, want, hear, need
6. covid, anti, just, tell, say, refuse, vaxxer, afraid, reason, people, stop, right, make, job, stupid
7. covid, year, trust, chance, inject, month, government, test, old, develop, cold, make, research, come
8. effect, know, long, term, covid, dna, affect, change, people, bad, rush, chance, unknown, study, test
9. people, covid, die, need, think, just, kill, family, care, damn, believe, say, real, death, chance
10. covid, force, try, reaction, people, bad, severe, look, allergic, medical, receive, say, fine, pay

Positive behavioral intentions

1. covid, people, want, just, think, say, know, mask, wear, make, really, ask, scare, right
2. covid, want, need, look, tomorrow, let, know, life, forward, ready, dose, morning, normal, receive, volunteer
3. covid, wait, long, turn, effect, line, term, finally, eat, excited, worried, afraid, use, drink, polio
4. just, dose, covid, second, got, day, effect, symptom, receive, fever, ache, hour, experience, headache, body
5. flu, covid, shot, year, time, bad, shoot, sick, immune, just, need, month, make, think, doctor
6. covid, arm, sore, sign, just, feel, today, hour, little, hurt, yesterday, injection, far, nervous, appointment
7. work, covid, home, thank, stay, patient, hospital, help, safe, care, protect, family, trial, receive, vaccinate
8. covid, risk, immune, die, people, virus, chance, high, know, need, vaccinate, healthy, live, just, catch
9. covid, today, hope, mom, test, dose, happy, able, soon, dad, positive, receive, good, grateful

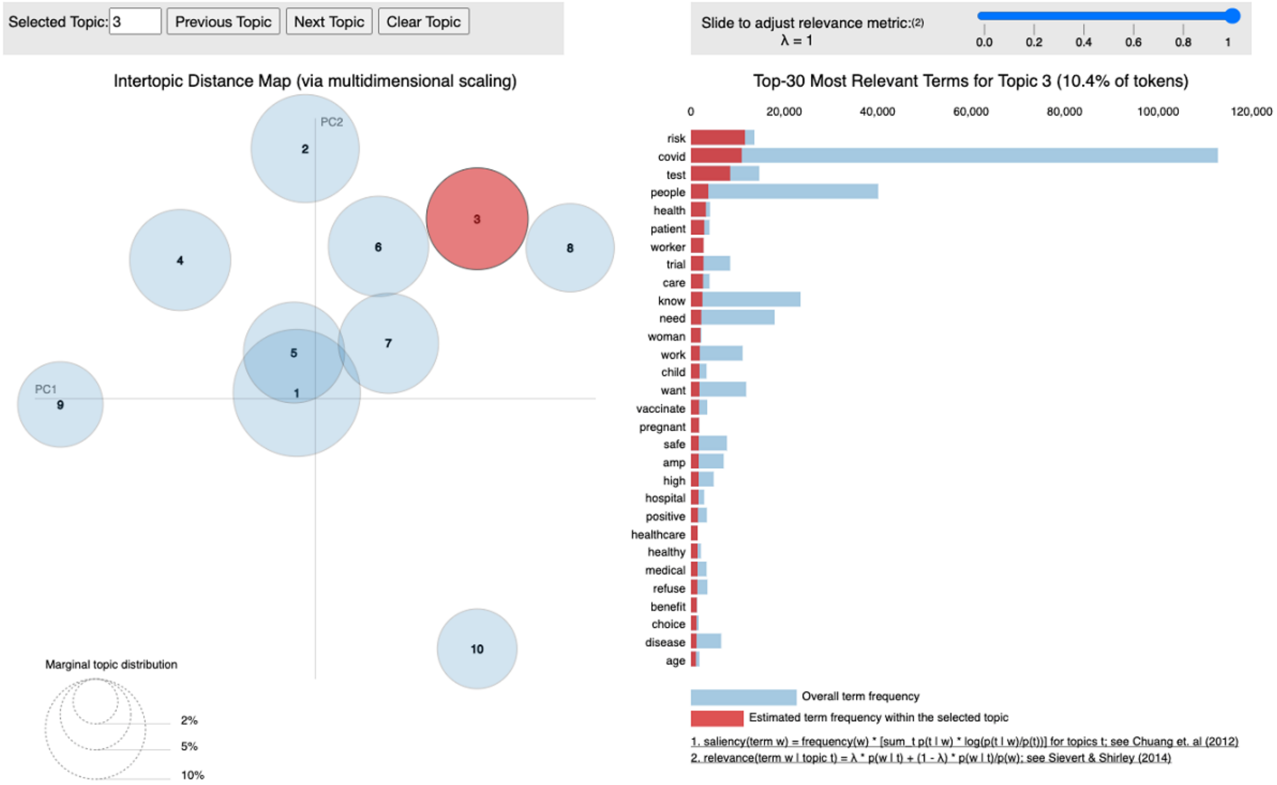
10. feel, covid, day, week, fine, great, make, shit, ago, better, worker, body, job, good, healthcare

Attitudes

Ten topics were extracted among the tweets that contained negative attitudes. The interactive display interface of pyLDAvis is shown in Figure 4. The left panel shows the distribution of topics, and we could choose the topic we wanted to analyze by clicking on the bubble (eg, topic 3 highlighted in Figure 4), while the right panel lists the top 30 relevant terms and their weights contributing to the selected topic. Some important keywords contained in topic 3 were “risk,” “test,” “child,” “safe,” “pregnant,” “disease,” and “age.” Topic 3 summarized that users with negative attitudes were concerned about the safety issues of the COVID-19 vaccines, especially about the risks for certain populations such as children, pregnant women, and patients with immune diseases. Other topics reflected concerns about unknown side effects (topic 4) and rushing the development process (topic 5). Some users even questioned the existence of COVID-19 or COVID-19 vaccines and indicated a lack of trust in the government or scientists (topic 8). In addition, some users feared that the virus mutation would render the vaccine ineffective (topic 7) and thus had negative attitudes toward vaccines.

For tweets containing positive attitudes, in a dominant topic (topic 3), relevant key terms included “safe,” “stay,” “end,” pandemic,” “news,” “effective,” “trial,” “continue,” and “hope.” This indicates that some positive attitudes might be derived from news of effective trial results and some users hoped that COVID-19 vaccines could end the pandemic. Relevant terms for topic 4 were “hope,” “normal,” “life,” “return,” “start,” “new,” “world,” and “great.” Tweets in topic 4 showed that some users expressed positive attitudes toward vaccines because of the desire to return to a normal life.

Figure 4. PyLDAvis visualization highlighting the top 30 relevant keywords for a topic found in the tweets that contained negative attitudes toward COVID-19 vaccines.


Selected Topic: 3
Previous Topic
Next Topic
Clear Topic
Slide to adjust relevance metric:(2)
λ = 1
0.0
0.2
0.4
0.6
0.8
1
Intertopic Distance Map (via multidimensional scaling)
PC2
PC1
1
2
3
4
5
6
7
8
9
10
Marginal topic distribution
2%
5%
10%
Top-30 Most Relevant Terms for Topic 3 (10.4% of tokens)
0
20,000
40,000
60,000
80,000
100,000
120,000
risk
covid
test
people
health
patient
worker
trial
care
know
need
woman
work
child
want
vaccinate
pregnant
safe
amp
high
hospital
positive
healthcare
healthy
medical
refuse
benefit
choice
disease
age
Overall term frequency
Estimated term frequency within the selected topic
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

Behavioral Intentions

For tweets containing negative behavioral intentions, topics 8 and 10 clustered independently; however, other topics showed some degree of mutual inclusiveness, indicating that similarities existed in those topics. Key terms for topic 8 were “effect,” “know,” “long,” “term,” “DNA,” “unknown,” and “rush.” This topic reflected that some users’ negative behavioral intentions came from the concerns of the long-term and unknown side effects of COVID-19 vaccines. As another unique topic, the most relevant terms for topic 10 were “force,” “reaction,” “bad,” “allergic,” “pay,” “adverse,” and “government.” This analysis highlighted that some users mentioned that they would not take the vaccine if it was forced on them by the government. Others worried about the adverse reactions to the COVID-19 vaccines. Some users compared COVID-19 to influenza and mentioned that because they had not previously been vaccinated against influenza, there was also no need to vaccinate against a disease they mistakenly thought had the same low lethality (topic 2). Other users reported that their immune system could naturally help them fight the virus.

For tweets containing positive behavioral intentions, mutual inclusivity existed among topics 1-4 and between topics 9 and 10. Other topics clustered independently. In topic 8, the keywords were “risk,” “immune,” “healthy,” “antibody,” and “immunity.” In this topic, users would like to become immune to the virus causing COVID-19 and stay healthy by being vaccinated.

Discussion

Principal Findings

In this study, we provided an annotated dataset with 5000 COVID-19 vaccine–related tweets with labels supporting three classification tasks (opinions, attitudes, and behavioral intentions). We assessed that transfer learning could be used to analyze COVID-19 vaccine content tweets and proved that they outperformed common machine learning models. We analyzed the temporal trends and topics in the COVID-19 vaccine–related tweets posted over a 3-month period (from November 1, 2020 to January 31, 2021). The prevalence of tweets containing positive behavioral intentions increased over time. The word clouds and the LDA analysis proved to be efficient tools to understand topics for tweets in each category.

Transfer learning is now widely used to analyze social media content. Some researchers have applied transfer learning with datasets of tweets related to COVID-19 [36-38] rather than focusing on tweets related to the vaccines developed for this disease. Researchers have analyzed tweets related to other vaccines such as HPV vaccines [25]. However, few studies have annotated tweets containing content about COVID-19 vaccines or developed models to understand public perceptions on COVID-19 vaccines from social media. For example, Levy et al [36] applied cross-lingual transfer learning to model COVID-19 outbreak patterns in one country, and then utilized the model to predict the spread of the disease in another country with a strong Spearman correlation (0.850). A classification model based on transfer learning developed by Spangher et al [37] was able to categorize policy announcements of COVID-19 using event extraction, with an F1 score of 0.770. To identify informative tweets related to COVID-19, Tasneem et al [38] proposed a unified architecture to combine transfer learning with hand-crafted features, achieving an F1 score of 0.820. Du et al [25] used deep learning models to categorize HPV vaccine–related tweets with constructs in the health belief model and theory of planned behavior models, and obtained F1 scores ranging from 0.681 to 0.942. Our study is the first to apply transfer learning models to analyze the public’s attitudes and behavioral intentions toward COVID-19 vaccines. Our model also achieved good performance, with F1 scores ranging from 0.579 to 0.792. In addition, we provided an annotated dataset with 5000 tweets, each labeled according to whether the tweet contained users’ opinions, attitudes, or behavioral intentions on COVID-19 vaccines. This dataset can be used for further research on social media content related to the COVID-19 vaccines.

Several researchers have applied the Valence Aware Dictionary and Sentiment Reasoner (VADER) tool [39,40], machine learning [41], and deep learning [42] to perform sentiment analysis on COVID-19–related tweets. Chandrasekaran et al [39] and Yin et al [40] employed the VADER tool to calculate the polarity of sentiment in COVID-19–related tweets posted in the first half of 2020. Both of these studies reported that the proportion of positive tweets was higher than that of negative tweets in general. However, Chandrasekaran et al [39] determined that negative tweets were dominant in the themes of symptoms and spread in cases. Li et al [42] used deep learning to identify fear and sadness emotions mentioned in COVID-19–related tweets to analyze the public’s mental health status, and reported area under the receiver operating characteristic curve values ranging from 0.681 to 0.739. Chakraborty et al [41] developed machine learning models with Gaussian membership function–based fuzzy rules to classify sentiment in COVID-19–related tweets, obtaining accuracy values ranging from 0.526 to 0.814. Although these previous studies have classified sentiment in COVID-19–related tweets, our study differs with respect to the task of classifying attitudes toward COVID-19 vaccines. We not only focused on the sentiment of tweets but also simultaneously examined whether the object of the sentiment was the COVID-19 vaccine. During annotation, we noticed that some tweets contained positive words used to describe what would happen after the vaccine rollout but also stated negative attitudes toward the vaccine itself, such as lack of trust and rushing.

Temporal analysis and topic modeling provide an efficient approach to monitor public perceptions of the COVID-19 vaccines on social media platforms. The following events could explain the significant increase in the prevalence of positive behavioral intentions in mid-December. For example, the FDA issued Pfizer-BioNTech COVID-19 vaccines on December 11, 2020, turning the vaccines from a hypothetical situation into a reality. The United States launched its rollout to high-risk health care facilities on December 14, 2020. A large number of health care workers and influential figures such as Joe Biden received COVID-19 vaccines to increase public confidence. This also suggests that more people might be willing to be vaccinated after successful vaccine development and a large-scale rollout. Indeed, social influence has been shown to positively affect the acceptance rate [43]. At the same time, this increase in positive behavioral intentions could also generate a positive social influence, which could lead to a higher vaccine acceptance rate. Therefore, the low acceptance rate of COVID-19 vaccines reported in the surveys conducted prior to December 2020 might not accurately reflect the current situation. Researchers should consider resurveying the public’s intention to receive the vaccination. Key terms identified in topic modeling could provide the needed guidance to design or optimize vaccine promotion interventions (eg, education materials). COVID-19 vaccine promotion strategies need to solve concerns on side effects and long-term safety issues, virus mutation, and the difference between COVID-19 and the flu. Moreover, promotion strategies should highlight the chance to return to normal life and stay healthy after being vaccinated for COVID-19.

Limitations

This study has several limitations. First, users of the Twitter platform are not representative of the entire public. The Twitter platform is usually considered to gather more antivaccinators and spread misinformation. This group of users is the main subgroup of the population with sentiments of vaccine hesitancy and should therefore be one of the main targets to receive vaccine education. Compared to other populations, they tend to question vaccines from specific perspectives such as the presence of microchips in vaccines [44] and the use of human embryos in the process of developing vaccines. Understanding their perceptions is a necessary step to tailor vaccine promotion education materials, which would provide a better chance of effectively changing their behavior. Second, some topics extracted from the topic modeling might be difficult to infer accurately using relevant terms. In addition, given the complex situation of behavioral intentions toward COVID-19 vaccines mentioned in the tweets, further qualitative studies (eg, content analysis) combined with theoretical models are needed to understand why some people will not get the vaccines from psychological aspects. Third, we applied the “BERT-base-uncase” as the language model. Recently, researchers have developed a transformer-based model COVID-Twitter-BERT (CT-BERT) model, which was pretrained on COVID-19–related tweets [45], and they also expected to obtain performance gains when applying the CT-BERT model on health care content tweets. The impact of using the CT-BERT model on our classification tasks is unknown. Fourth, the annotated corpus included 5000 tweets. If more annotated data could be collected, the performance of the model might be improved.

For future work, we will perform a theory-based content analysis to gain insight into the reasons that led to the changes in behavioral intentions we noted in the temporal analysis. Using the transfer learning model in this study, researchers can automatically collect tweets containing COVID-19 vaccine–related behavioral intentions and systematically analyze the data through a theoretical model (eg, Capability, Opportunity, Motivation, Behavior model [12,46]) to promote timely promotion strategies. In addition, researchers can extract individual characteristics from the user profile and perform statistical analysis to determine the relationship between individual characteristics and their behavioral intention toward COVID-19 vaccines.

Conclusion

In this study, we presented an annotated corpus of 5000 tweets and analyzed the potential to use transfer learning with a pretrained BERT model to automatically identify public opinions, behavioral intentions, and attitudes toward COVID-19 vaccines from social media. We demonstrated that transfer learning models outperformed traditional machine learning models in general. In addition, we explored the temporal trends of the public’s change in attitudes and behavioral intentions on a larger dataset with 2,678,372 tweets from November 1, 2020 to January 31, 2021. We found that the LDA technique is useful to extract topics from identified tweets. Overall, we provided an automatic method to analyze the public’s understanding of COVID-19 vaccines from real-time data, which could be used to tailor education programs and other interventions to promote COVID-19 vaccine acceptance urgently.

Acknowledgments

This work was supported by Sichuan Science and Technology Program (grant number 2020YFS0162).

Authors' Contributions

JLiu and SL conceived the study. SL, JLiu, and JLi performed the analysis, interpreted the results, and drafted the manuscript. All authors revised the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. WHO Coronavirus Disease (COVID-19) Dashboard. World Health Organization. 2020. URL: https://covid19.who.int/ [accessed 2021-01-31]
2. Chou WS, Budenz A. Considering emotion in COVID-19 vaccine communication: addressing vaccine hesitancy and fostering vaccine confidence. Health Commun 2020 Dec;35(14):1718-1722. [doi: 10.1080/10410236.2020.1838096] [Medline: 33124475]
3. Liu J, Liu S. The management of coronavirus disease 2019 (COVID-19). J Med Virol 2020 Sep;92(9):1484-1490 [FREE Full text] [doi: 10.1002/jmv.25965] [Medline: 32369222]
4. Liu J, Liu S, Wei H, Yang X. Epidemiology, clinical characteristics of the first cases of COVID-19. Eur J Clin Invest 2020 Oct;50(10):e13364. [doi: 10.1111/eci.13364] [Medline: 32725884]
5. Mahase E. Covid-19: What do we know about the late stage vaccine candidates? BMJ 2020 Nov 24;371:m4576. [doi: 10.1136/bmj.m4576] [Medline: 33234507]
6. Kreps S, Prasad S, Brownstein JS, Hswen Y, Garibaldi BT, Zhang B, et al. Factors associated with US adults' likelihood of accepting COVID-19 vaccination. JAMA Netw Open 2020 Oct 01;3(10):e2025594 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.25594] [Medline: 33079199]

7. Chodosh S. Why only half of Americans say they would get a COVID-19 vaccine. Popular Science. 2020 Jun 01. URL: https://www.popsci.com/story/health/covid-19-vaccine-poll/ [accessed 2021-01-31]
8. D'souza G, Dowdy D. What is herd immunity and how can we achieve it with COVID-19? Johns Hopkins Bloomberg School of Public Health. 2020. URL: https://www.jhsph.edu/covid-19/articles/achieving-herd-immunity-with-covid19.html [accessed 2021-01-31]
9. Kwok KO, Li K, Wei WI, Tang A, Wong SYS, Lee SS. Editor's Choice: Influenza vaccine uptake, COVID-19 vaccination intention and vaccine hesitancy among nurses: A survey. Int J Nurs Stud 2021 Feb;114:103854 [FREE Full text] [doi: 10.1016/j.ijnurstu.2020.103854] [Medline: 33326864]
10. 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]
11. Puri N, Coomes EA, Haghbayan H, Gunaratne K. Social media and vaccine hesitancy: new updates for the era of COVID-19 and globalized infectious diseases. Hum Vaccin Immunother 2020 Nov 01;16(11):2586-2593 [FREE Full text] [doi: 10.1080/21645515.2020.1780846] [Medline: 32693678]
12. Liu S, Liu J. Understanding behavioral intentions toward COVID-19 vaccines: theory-based content analysis of tweets. J Med Internet Res 2021 May 12;23(5):e28118 [FREE Full text] [doi: 10.2196/28118] [Medline: 33939625]
13. Tyson A, Johnson C, Funk C. U.S. public now divided over whether to get COVID-19 vaccine. Pew Research Center. 2020. URL: https://www.pewresearch.org/science/ [accessed 2021-01-31]
14. COCONEL Group. A future vaccination campaign against COVID-19 at risk of vaccine hesitancy and politicisation. Lancet Infect Dis 2020 Jul;20(7):769-770 [FREE Full text] [doi: 10.1016/S1473-3099(20)30426-6] [Medline: 32445713]
15. 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]
16. Brennen J, Simon F, Howard P, Nielsen R. Types, sources, and claims of COVID-19 misinformation key findings. Reuters Institute, University of Oxford. URL: https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation [accessed 2021-01-30]
17. Chan TM, Dzara K, Dimeo SP, Bhalerao A, Maggio LA. Social media in knowledge translation and education for physicians and trainees: a scoping review. Perspect Med Educ 2020 Feb;9(1):20-30 [FREE Full text] [doi: 10.1007/s40037-019-00542-7] [Medline: 31834598]
18. Lwin MO, Lu J, Sheldenkar A, Schulz PJ. Strategic uses of Facebook in Zika outbreak communication: implications for the crisis and emergency risk communication model. Int J Environ Res Public Health 2018 Sep 10;15(9):1974 [FREE Full text] [doi: 10.3390/ijerph15091974] [Medline: 30201929]
19. Zhang EX, Yang Y, Di Shang R, Simons JJP, Quek BK, Yin XF, et al. Leveraging social networking sites for disease surveillance and public sensing: the case of the 2013 avian influenza A(H7N9) outbreak in China. Western Pac Surveill Response J 2015;6(2):66-72 [FREE Full text] [doi: 10.5365/WPSAR.2015.6.1.013] [Medline: 26306219]
20. Mourad A, Srour A, Harmanani H, Jenainati C, Arafeh M. Critical impact of social networks infodemic on defeating coronavirus COVID-19 pandemic: Twitter-based study and research directions. IEEE Trans Netw Serv Manage 2020 Dec 01;17(4):2145-2155. [doi: 10.1109/tnsm.2020.3031034]
21. Mavragani A. Tracking COVID-19 in Europe: infodemiology approach. JMIR Public Health Surveill 2020 Apr 20;6(2):e18941 [FREE Full text] [doi: 10.2196/18941] [Medline: 32250957]
22. Li J, Xu Q, Cuomo R, Purushothaman V, Mackey T. Data mining and content analysis of the Chinese social media platform Weibo during the early COVID-19 outbreak: retrospective observational infoveillance study. JMIR Public Health Surveill 2020 Apr 21;6(2):e18700 [FREE Full text] [doi: 10.2196/18700] [Medline: 32293582]
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. Fadda M, Galimberti E, Fiordelli M, Romanò L, Zanetti A, Schulz PJ. Effectiveness of a smartphone app to increase parents' knowledge and empowerment in the MMR vaccination decision: A randomized controlled trial. Hum Vaccin Immunother 2017 Nov 02;13(11):2512-2521 [FREE Full text] [doi: 10.1080/21645515.2017.1360456] [Medline: 29125783]
25. Du J, Luo C, Shegog R, Bian J, Cunningham RM, Boom JA, et al. Use of deep learning to analyze social media discussions about the human papillomavirus vaccine. JAMA Netw Open 2020 Nov 02;3(11):e2022025 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.22025] [Medline: 33185676]
26. Liu J, Wu J, Liu S, Li M, Hu K, Li K. Predicting mortality of patients with acute kidney injury in the ICU using XGBoost model. PLoS One 2021;16(2):e0246306 [FREE Full text] [doi: 10.1371/journal.pone.0246306] [Medline: 33539390]
27. Liu S, Ferraro J, Gundlapalli AV, Chapman W, Bucher B. Detection of healthcare-associated infections using electronic health record data. 2018 Nov Presented at: AMIA 2018 Annual Symposium; November 3-7, 2018; San Francisco, CA.
28. Shi J, Liu S, Pruitt LCC, Luppens CL, Ferraro JP, Gundlapalli AV, et al. Using natural language processing to improve EHR structured data-based surgical site infection surveillance. AMIA Annu Symp Proc 2019;2019:794-803 [FREE Full text] [Medline: 32308875]

29. Devlin J, Chang M, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. 2019 Jun Presented at: NAACL-HLT 2019; 2019; Minneapolis, MN p. 4171-4186.
30. Demsar J. Statistical comparisons of classifiers over multiple data sets. J Mach Learn Res 2006 Dec 01;7(1):1-30 [FREE Full text]
31. Dickey DA, Fuller WA. Distribution of the estimators for autoregressive time series with a unit root. J Am Stat Assoc 1979 Jun;74(366a):427-431. [doi: 10.1080/01621459.1979.10482531]
32. Hidayatullah A, Pembrani E, Kurniawan W, Akbar G, Pranata R. Twitter topic modeling on football news. 2018 Apr Presented at: 2018 3rd International Conference on Computer and Communication Systems; 2018; Nagoya, Japan p. 467-471.
33. Blei DM. Probabilistic topic models. Communications of the ACM 2012;55(4):77-84.
34. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. J Machine Learn Res 2003 Mar 01;3:993-1022. [doi: 10.3156/jsoft.24.4_160_1]
35. Sievert C, Shirley K. LDAvis: A method for visualizing and interpreting topics. 2014 Jun 27 Presented at: Workshop on Interactive Language Learning, Visualization, and Interfaces; 2014; Baltimore, MD p. 63-70.
36. Levy S, Wang W. Cross-lingual transfer learning for COVID-19 outbreak alignment. arXiv. URL: http://arxiv.org/abs/2006.03202 [accessed 2021-02-11]
37. Spangher A, Peng N, May J, Ferrara E. Enabling low-resource transfer learning across COVID-19 corpora by combining event-extraction and co-training. : Association for Computational Linguistics; 2020 Presented at: ACL 2020 Work NLP-COVID 2020; July 2020; online URL: https://aclanthology.org/2020.nlpcovid19-acl.4
38. Tasneem F, Naim J, Tasnia R, Hossain T, Chy A. CSECU-DSG at WNUT-2020 Task 2: Exploiting ensemble of transfer learning and hand-crafted features for identification of informative COVID-19 English tweets. 2020 Nov 19 Presented at: 2020 EMNLP Workshop W-NUT: The Sixth Workshop on Noisy User-generated Text; November 19, 2020; Online p. 394-398.
39. Chandrasekaran R, Mehta V, Valkunde T, Moustakas E. Topics, trends, and sentiments of tweets about the COVID-19 pandemic: temporal infoveillance study. J Med Internet Res 2020 Oct 23;22(10):e22624 [FREE Full text] [doi: 10.2196/22624] [Medline: 33006937]
40. Yin H, Yang S, Li J. Detecting topic and sentiment dynamics due to COVID-19 pandemic using social media. arXiv. 2020 Jul 05. URL: http://arxiv.org/abs/2007.02304 [accessed 2021-02-11]
41. Chakraborty K, Bhatia S, Bhattacharyya S, Platos J, Bag R, Hassanien AE. Sentiment analysis of COVID-19 tweets by deep learning classifiers-a study to show how popularity is affecting accuracy in social media. Appl Soft Comput 2020 Dec;97:106754 [FREE Full text] [doi: 10.1016/j.asoc.2020.106754] [Medline: 33013254]
42. Li I, Li Y, Li T, Alvarez-Napagao S, Garcia D. What are we depressed about when we talk about COVID19: mental health analysis on tweets using natural language processing. arXiv. 2020 Jun 08. URL: https://arxiv.org/abs/2004.10899 [accessed 2021-02-11]
43. Liu S, Reese TJ, Kawamoto K, Del Fiol G, Weir C. A systematic review of theoretical constructs in CDS literature. BMC Med Inform Decis Mak 2021 Mar 17;21(1):102 [FREE Full text] [doi: 10.1186/s12911-021-01465-2] [Medline: 33731089]
44. Evanega S, Lynas M, Adams J, Smolenyak K. Coronavirus misinformation: quantifying sources and themes in the COVID-19 "infodemic". Alliance for Science. URL: https://allianceforscience.cornell.edu/wp-content/uploads/2020/10/Evanega-et-al-Coronavirus-misinformation-submitted_07_23_20-1.pdf [accessed 2021-02-11]
45. Müller M, Salathé M, Kummervold P. COVID-Twitter-BERT: a natural language processing model to analyse COVID-19 content on Twitter. arXiv. 2020 May 15. URL: https://arxiv.org/abs/2005.07503 [accessed 2021-01-28]
46. 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]

Abbreviations

ADF: Augmented Dickey-Fuller
BERT: bidirectional encoder representations from transformers
CT-BERT: COVID-Twitter-bidirectional encoder representations from transformers
FDA: Food and Drug Association
HPV: human papillomavirus vaccine
LDA: latent Dirichlet allocation
VADER: Valence Aware Dictionary and Sentiment Reasoner
WHO: World Health Organization

Edited by C Basch; submitted 07.05.21; peer-reviewed by J Lei, C Huang, R Poluru; comments to author 09.06.21; revised version received 28.06.21; accepted 11.07.21; published 10.08.21.

Please cite as:
Liu S, Li J, Liu J
Leveraging Transfer Learning to Analyze Opinions, Attitudes, and Behavioral Intentions Toward COVID-19 Vaccines: Social Media Content and Temporal Analysis
J Med Internet Res 2021;23(8):e30251
URL: https://www.jmir.org/2021/8/e30251
doi:10.2196/30251
PMID:34254942

©Siru Liu, Jili Li, Jialin Liu. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Long-term Effects of the COVID-19 Pandemic on Public Sentiments in Mainland China: Sentiment Analysis of Social Media Posts

Hao Tan[1,2], DPhil; Sheng-Lan Peng[1,2], ME; Chun-Peng Zhu[1,2], BA; Zuo You[1,2], ME; Ming-Cheng Miao[3], BA; Shu-Guang Kuai[3,4], DPhil

[1]School of Design, Hunan University, Changsha, China

[2]Emergency Science Joint Research Center, Hunan University, Changsha, China

[3]Shanghai Key Laboratory of Mental Health and Psychological Crisis Intervention, Key Laboratory of Brain Functional Genomics (Ministry of Education), The Institute of Brain and Education Innovation, School of Psychology and Cognitive Science, East China Normal University, Shanghai, China

[4]NYU-ECNU Institute of Brain and Cognitive Science, New York University Shanghai, Shanghai, China

Corresponding Author:
Shu-Guang Kuai, DPhil
Shanghai Key Laboratory of Mental Health and Psychological Crisis Intervention
Key Laboratory of Brain Functional Genomics (Ministry of Education), The Institute of Brain and Education Innovation
School of Psychology and Cognitive Science, East China Normal University
Room 309, Old Library Building
No. 3663, North Zhongshan Road
Shanghai, 200062
China
Phone: 86 15618582866
Email: shuguang.kuai@gmail.com

Abstract

Background: The COVID-19 outbreak has induced negative emotions among people. These emotions are expressed by the public on social media and are rapidly spread across the internet, which could cause high levels of panic among the public. Understanding the changes in public sentiment on social media during the pandemic can provide valuable information for developing appropriate policies to reduce the negative impact of the pandemic on the public. Previous studies have consistently shown that the COVID-19 outbreak has had a devastating negative impact on public sentiment. However, it remains unclear whether there has been a variation in the public sentiment during the recovery phase of the pandemic.

Objective: In this study, we aim to determine the impact of the COVID-19 pandemic in mainland China by continuously tracking public sentiment on social media throughout 2020.

Methods: We collected 64,723,242 posts from Sina Weibo, China's largest social media platform, and conducted a sentiment analysis based on natural language processing to analyze the emotions reflected in these posts.

Results: We found that the COVID-19 pandemic not only affected public sentiment on social media during the initial outbreak but also induced long-term negative effects even in the recovery period. These long-term negative effects were no longer correlated with the number of new confirmed COVID-19 cases both locally and nationwide during the recovery period, and they were not attributed to the postpandemic economic recession.

Conclusions: The COVID-19 pandemic induced long-term negative effects on public sentiment in mainland China even as the country recovered from the pandemic. Our study findings remind public health and government administrators of the need to pay attention to public mental health even once the pandemic has concluded.

(J Med Internet Res 2021;23(8):e29150) doi:10.2196/29150

KEYWORDS

COVID-19; emotional trauma; public sentiment on social media; long-term effect

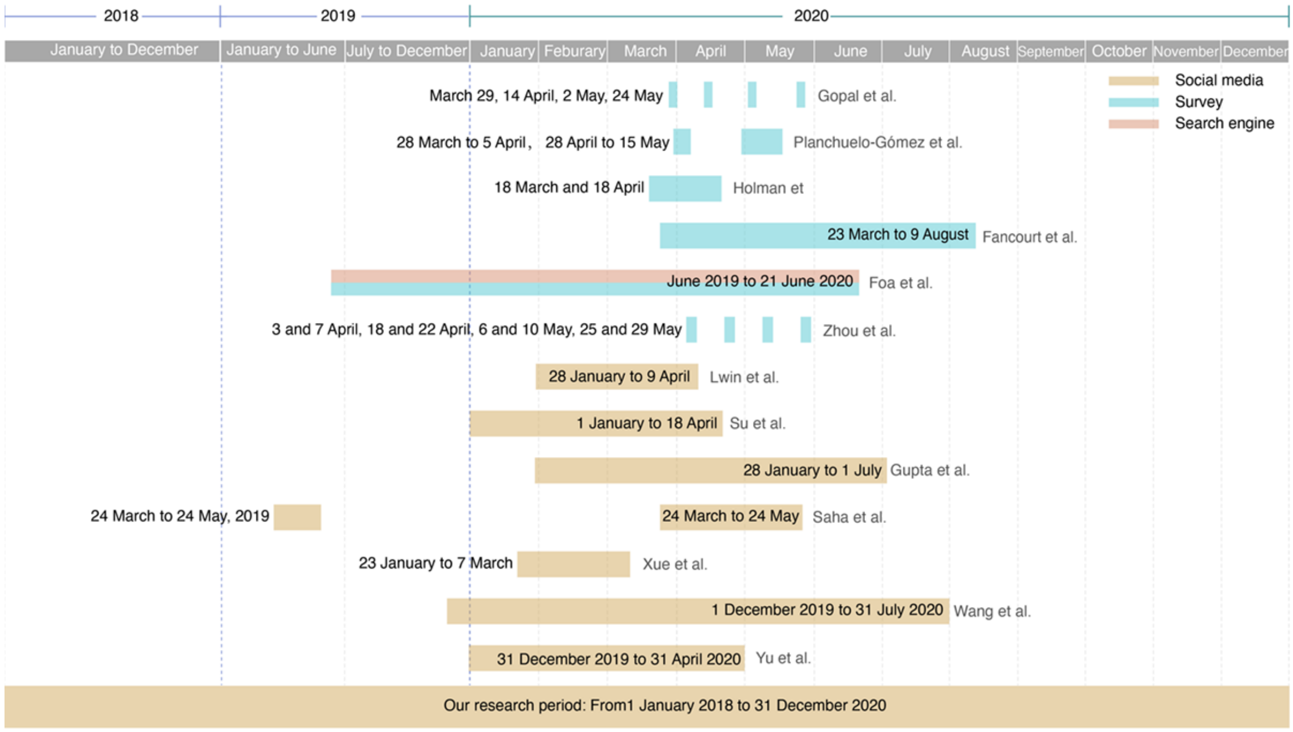
Introduction

The COVID-19 outbreak has spread rapidly across the world, leading to a global pandemic. At the end of 2020, more than 82,662,478 confirmed cases of infection and at least 1,872,802 deaths were reported globally [1]. To prevent the spread of the pandemic, many countries imposed different levels of restrictions in their administrations, such as enforcing statewide lockdowns, home isolation, and travel bans. COVID-19, similar to previous infectious disease outbreaks such as SARS (severe acute respiratory syndrome) in 2003 and Ebola virus disease in 2014, has not only threatened the physical health of the public but also imposed a wide range of negative emotions, including fear, depression, and panic disorder [2-4]. Such negative emotions could be harmful to the public's mental health and even trigger social unrest [5]. Therefore, understanding how the pandemic affects public sentiment can provide valuable information for policymakers, government administrators, and mental health service providers.

Since the onset of the pandemic, researchers have conducted online and offline surveys to assess the public’s mental health. These surveys have consistently shown that the COVID-19 outbreak has had a devastating negative impact on the public’s mental health [6-9]. Meanwhile, with the growing popularity of the internet, people have extensively expressed their emotions through web-based platforms such as microblogs. The development of natural language processing allows us to identify and quantify people's emotional states by analyzing the content of their posts on the internet. The average emotional state of the entire community is defined as the *public sentiment*. Sentiment analysis of social media posts has been used in many previous studies and is believed to be a good indicator of public emotions in the internet era. Such sentiment analyses have indicated a significant increase in negative emotions and a decrease in positive emotions and life satisfaction [10,11]. Analyzing queries in search engines has also identified an increase in topics related to anxiety, negative thoughts, sleep disturbances, and even suicidal ideations [12,13]. Upon combining the evidence of studies by using different approaches, the negative impact of the outbreak on public mental health is unquestionable. However, it remains largely unknown how public sentiment has changed across various stages of the pandemic, such as the accelerating, decelerating, and “the new normal” stages.

Several tracking surveys have found an increase in negative emotional rating scores following the COVID-19 outbreak, as illustrated in Figure 1. These include Gopal et al in India [14], Planchuedlo-Gomez et al in Spain [15], Holman et al in the United States [7], among others. However, a literature search has yielded some contrasting findings. Fancourt et al [6] and Foa et al [12] found that negative emotional rating scores were the highest at the beginning of the outbreak and gradually decreased thereafter. Zhou et al [16] found a slight improvement in mood after the unlock phase of Wuhan, China, was initiated, by comparing residents’ mental health before and after the city’s reopening [16]. Moreover, studies using content analysis of social media and search engines showed the highest concerns and negative emotions to outbreak-related topics, and there was a gradual decrease in such negative emotions thereafter, as reported by Lwin et al [17], Su et al [11], Gupta et al [18], Saha et al [19], Xue et al [20], Wang et al [21], and Yu et al [22] (Figure 1).

Figure 1. A literature map summarizing published studies on tracking public mental health during the COVID-19 pandemic. The horizontal axis labels the tracking time and different shading colors represent varying approaches to assess mental health.



Most studies thus far have focused on the psychological impact of the pandemic in the first half of the year, and neither surveys nor social media sentiment analyses have assessed the public sentiment after August 2020 to date. The long-term effects of the pandemic on public sentiment, therefore, remain largely unknown. As mainland China was the first region hit by the COVID-19 pandemic, and it has been the first major economy to successfully recover from the pandemic shock [23,24], monitoring public sentiment on social media in mainland China at different stages of the pandemic would provide a comprehensive understanding of the pandemic's effect on public mental health. In this study, we tracked posts across Sina Weibo—the largest microblogging social media platform in China [25], which is analogous to Twitter worldwide. We analyzed the public sentiment reflected in Weibo posts across 2020 and identified interesting short- and long-term emotional trauma induced by the COVID-19 pandemic among Weibo users.

Methods

Collection of Weibo Posts

We used a web crawler technology to collect 64,723,242 Weibo posts from 26,895,593 accounts across 31 provinces, from January 1, 2018, to December 2020 (n=10,072,413 in 2018; n=21,652,707 in 2019; and n=32,998,122 in 2020). We tried to maintain uniform sampling across provinces to balance regional differences, with the number of posts published in each province ranging around 3000 posts per day in 2020, 2000 posts per day in 2019, and 1000 posts per day in 2018.

We calculated the number of Weibo posts per account. The data follow the Zipf distribution, which aligns with the results of Li et al [26]. A small number of users generated a very high number of posts; these are likely to be advertising accounts. We filtered out 0.01% of accounts with the highest number of posts and retained the accounts that were within 99.99% of the distribution of the Weibo post per capita. Moreover, accounts that were "verified" as those of stars, public figures, organizations, etc, were filtered out. Posts that appeared to be advertisements for marketing purposes were excluded as well. Additionally, we filtered out the posts that did not contain Chinese characters and those that exceeded 140 characters. The study was approved by the ethics committees of the East China Normal University and School of Design, Hunan University.

Calculation of Sentiment Values

We applied the Tencent natural language processing product [27], a professional Chinese sentiment analysis application programming interface (API), to analyze public sentiment on the internet [28]. The algorithm excluded numbers, punctuations, English characters, URL, hashtags, mentions, and emojis and then extracted Chinese characters, numbers, and punctuation for sentiment analysis. The algorithm generated a sentiment value score ranging from 0 (extremely negative mood) to 1.0 (strongly positive mood). We averaged mean sentiment values for all 31 provinces in mainland China.

Data of Reported COVID-19 Cases

We obtained the data on COVID-19 cases reported in mainland China from the National Health Commission of the People's Republic of China and the health commissions of municipal provinces [29]. We collected data regarding global cases of COVID-19 from the World Health Organization and Johns Hopkins Coronavirus Resource Center [1,30].

Data of Economic Indicators

We obtained data of economic indicators from the National Bureau of Statistics [31], including the consumer price index, unemployment rate in urban areas of mainland China, producer price index, and growth rate of gross domestic product (GDP). The first three indicators were calculated per month and the GDP growth rate was calculated per quarter.

Results

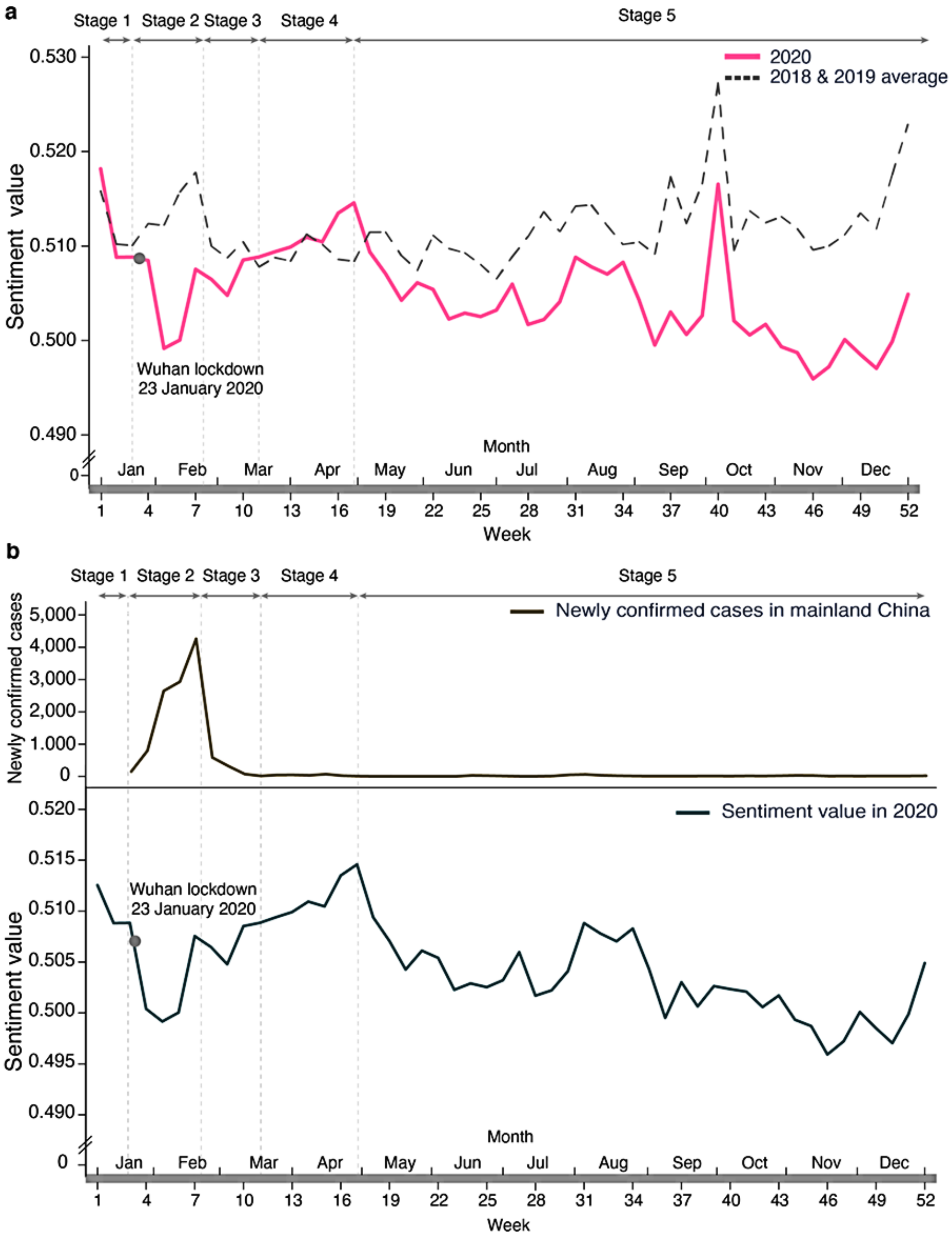
Social Media Public Sentiment Values Throughout 2020 and in the Previous Two Years

In week 3 (January 20, 2020), human-to-human transmission of the virus was confirmed and announced to the public. Thereafter, the city of Wuhan entered a lockdown on January 23, 2020 [32], attracting considerable public attention. As a consequence, social media public sentiment values rapidly decreased, hitting the lowest in week 5 (Figure 2a). The lowest sentiment value was reduced by more than 3.7% compared to the beginning of the year (week 1). The sentiment values bottomed out from week 6 and reached the peak at week 17 (Figure 2a). Thereafter, sentiment values surprisingly entered a long-term downward spiral until the end of the year, although the pandemic was well under control in mainland China. The lowest sentiment value (mean 0.496, SE 0.002) in the second half of the year was even lesser than that in week 5 (mean 0.499, SE 0.003), marking the lowest point after the COVID-19 outbreak in stage 2 (Figure 2a).

It is unclear whether the decline in sentiment values in the second half of 2020 presents a long-term effect of the pandemic or whether they are merely seasonal fluctuations. To answer this question, we analyzed public sentiment values from the internet in 2018 and 2019. Interestingly, the sentiment values in the previous two years did not demonstrate a downward trend in the second half of the year, indicating that the downward trend is probably not a result of common seasonal fluctuations (Figure 2a). It is worth noting that we observed three peaks of public sentiment values in week 1, week 7, and week 40, which were around the New Year, Spring Festival (ie, Chinese New Year), and the National Day, respectively, indicating the effects of holidays on public sentiment on the internet. The peak around the Spring Festival period observed in the previous two years did not occur in 2020. As the Chinese New Year was just after the COVID-19 outbreak, one might suspect that the negative effects of the pandemic were diluted by its holiday effects. To obtain a more accurate reading of the effect of the pandemic on public sentiment on the internet, it is necessary to exclude these holiday effects. Therefore, we estimated the holiday effects by using data of the previous two years. In particular, we computed the difference between sentiment values in the holiday week and those in the week before and after each holiday. Thus, the holiday effects were excluded from the sentiment values in 2020 for further analysis (Figure 2b).

According to the white paper published by the State Council Information Office of the People's Republic of China, titled “Fighting COVID-19 China in Action” [24], China divided its fight against the pandemic into five stages. The variations of sentiment values in 2020 were well aligned with these five stages of outbreak prevention. In stage 1 (weeks 0-3: *swift response to the public health emergency*), the sentiment values had been relatively stable given the pandemic had not yet triggered nationwide attention. In stage 2 (weeks 3-8: *initial progress in containing the virus*), due to public concern about the pandemic outbreak, the sentiment values rapidly declined and then rebounded quickly as the government brought the pandemic under control. In stage 3 (weeks 8-11: *newly confirmed domestic cases in mainland China decline to single digits*), sentiment values were restored to a stable level as the number of confirmed COVID-19 cases gradually decreased to single digits. In stage 4 (weeks 12-17: *Wuhan and Hubei—an initial victory in a critical battle*), sentiment values once again increased and even exceeded those in the same period in previous years. The sentiment values in stages 2-4 significantly correlated with the number of newly confirmed cases (Pearson correlation, *r*=–0.58, *P*=.02; see Figure 2b), indicating a strong correlation between public sentiment on social media and the severity of the pandemic.

Figure 2. Public sentiment values on the Chinese social media platform Weibo over the last 3 years. (a) Public sentiment values were plotted as a function of the week across the whole year (2020) or averaged values of the previous 2 years. (b) The number of newly confirmed COVID-19 cases per week for 2020 (upper panel) and sentiment values in 2020 excluding holiday effects (lower panel).

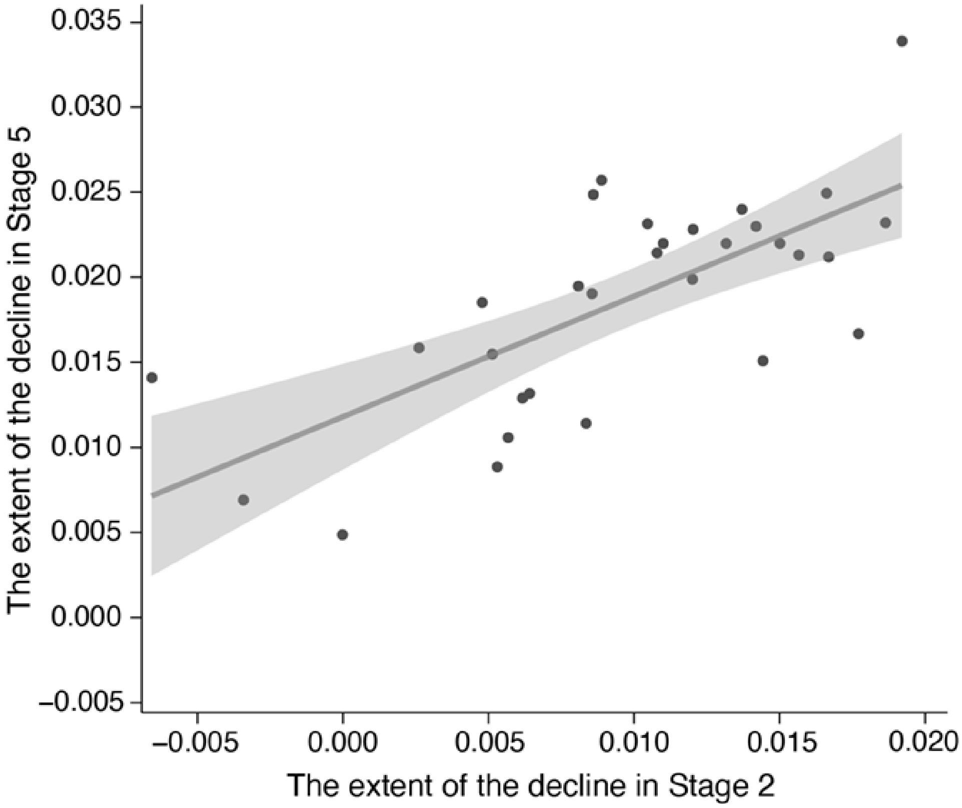


However, in stage 5 (week 18 to present: *ongoing prevention and control*), there was another unexpected decline in the public sentiment values even though the pandemic was well under control at the time. Interestingly, public sentiment values in this stage were no longer correlated with the number of newly confirmed COVID-19 cases (r=0.12, P=.50).

Relationship Between the Decline in Sentiment Values in Stages 2 and 5

In this study, we considered the underlying reasons for the decline in public sentiment on social media in stage 5. We hypothesized that the decline in sentiment values could be an after-effect of the decline in public sentiment in stage 2. Considering the different depths of the pandemic's severity across China, the extent of the decline in public sentiment values similarly varied across different provinces and municipalities. We reasoned that if there was a correlation between the extent of the decline in public sentiment values in stages 2 and 5, noting that it is likely that the decline in stage 5 reflects a long-term consequence of the outbreak in stage 2. To test the hypothesis, we calculated the difference of sentiment values between week 3 and 5 as the extent of the decline in stage 2 and the difference of sentiment values between weeks 17 and 46 as the extent of the decline in stage 5. We observed a significant positive correlation between the extent of the decline in stage 2 and that in stage 5 (r=0.71, P<.001), which supports our hypothesis that the decline in public sentiment values in stage 5 reflects the long-term emotional consequences induced by the initial COVID-19 outbreak (Figure 3).

Figure 3. A correlation between the declines observed in social media public sentiment values in stage 2 with those in stage 5. Black dots represent the extent of the decline of public sentiment values in each province in mainland China. The solid line represents the fitted linear curve of the extent of the decline in two stages (R^2=.50, P<.001, 95% CI 0.473-0.978).

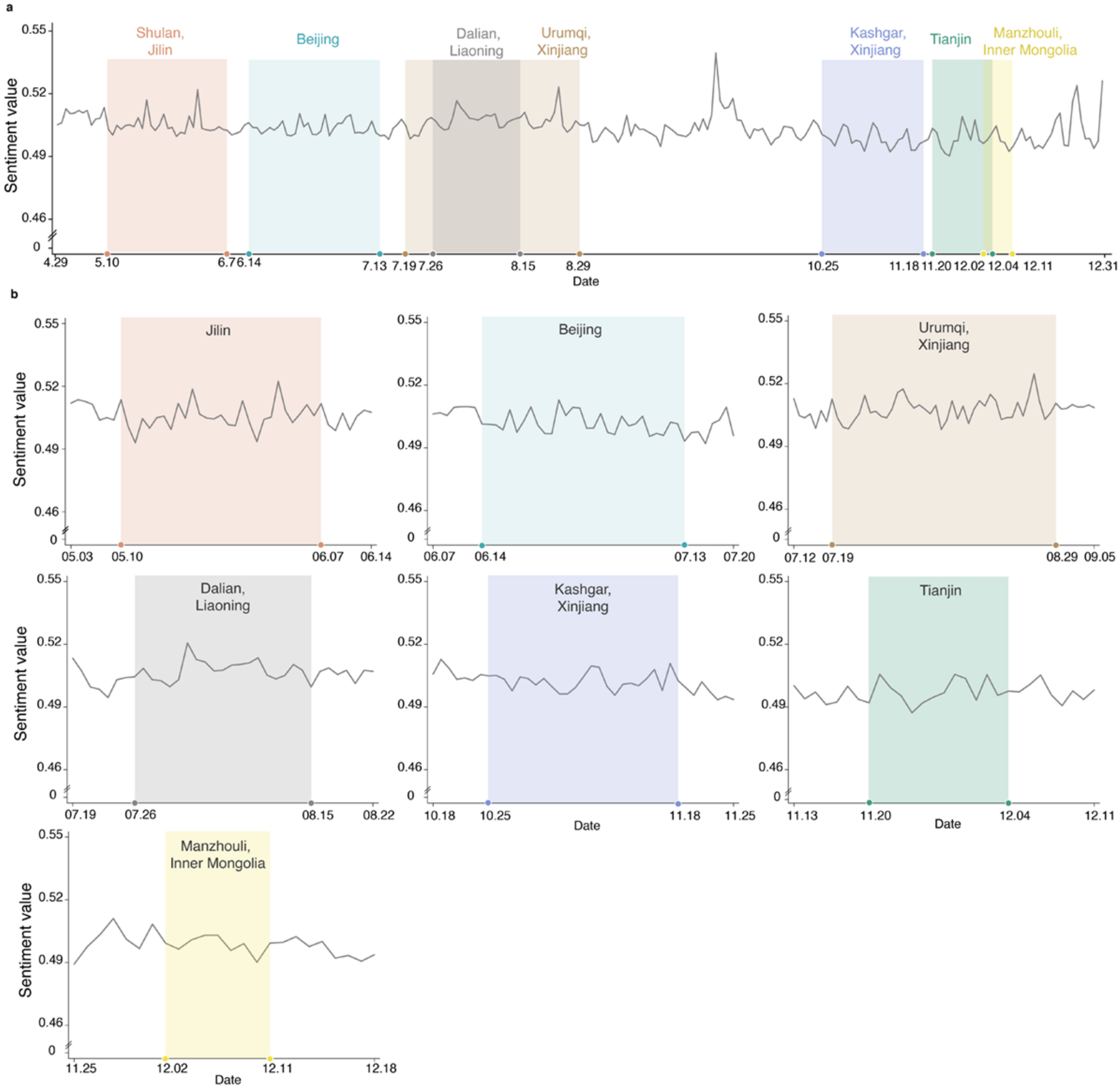


Alternative Causes Leading to the Decline in Public Sentiment on Social Media in Stage 5

Concerns About Upgrading Risk Assessment Level

Sporadic indigenous COVID-19 cases remained in stage 5. In response to the emergence of these local cases, governments upgraded the risk assessment level from *low* to *high* in Jilin (weeks 19-23), Beijing (weeks 24-28), Xinjiang (weeks 29-35), Liaoning (weeks 30-33), Xinjiang (weeks 43-47), Tianjin (weeks 47-49), and Inner Mongolia (weeks 49-50). Upgrading the risk assessment level may have induced negative emotions in the general public. To assess whether that was indeed the case, we defined average public sentiment values 1 week before and after the high-risk period as the baseline. We defined the rate of change by the difference between the average public sentiment values in the high-risk period and the baseline, then divided by the baseline. The rate of change was close to 0 either at the nationwide scale (Jilin: –0.19%, Beijing: 0.18%, Xinjiang: 0.94%, Liaoning: 0.61%, Xinjiang: –0.18%, Tianjin: 0.28%, Inner Mongolia: –0.21%; see Figure 4a) or local scales (Jilin: –0.13%, Beijing: –0.44%, Xinjiang: 0.17%, Liaoning: 0.57%, Xinjiang: 0.20%, Tianjin: 0.21%, Inner Mongolia: 0.05%; see Figure 4b).

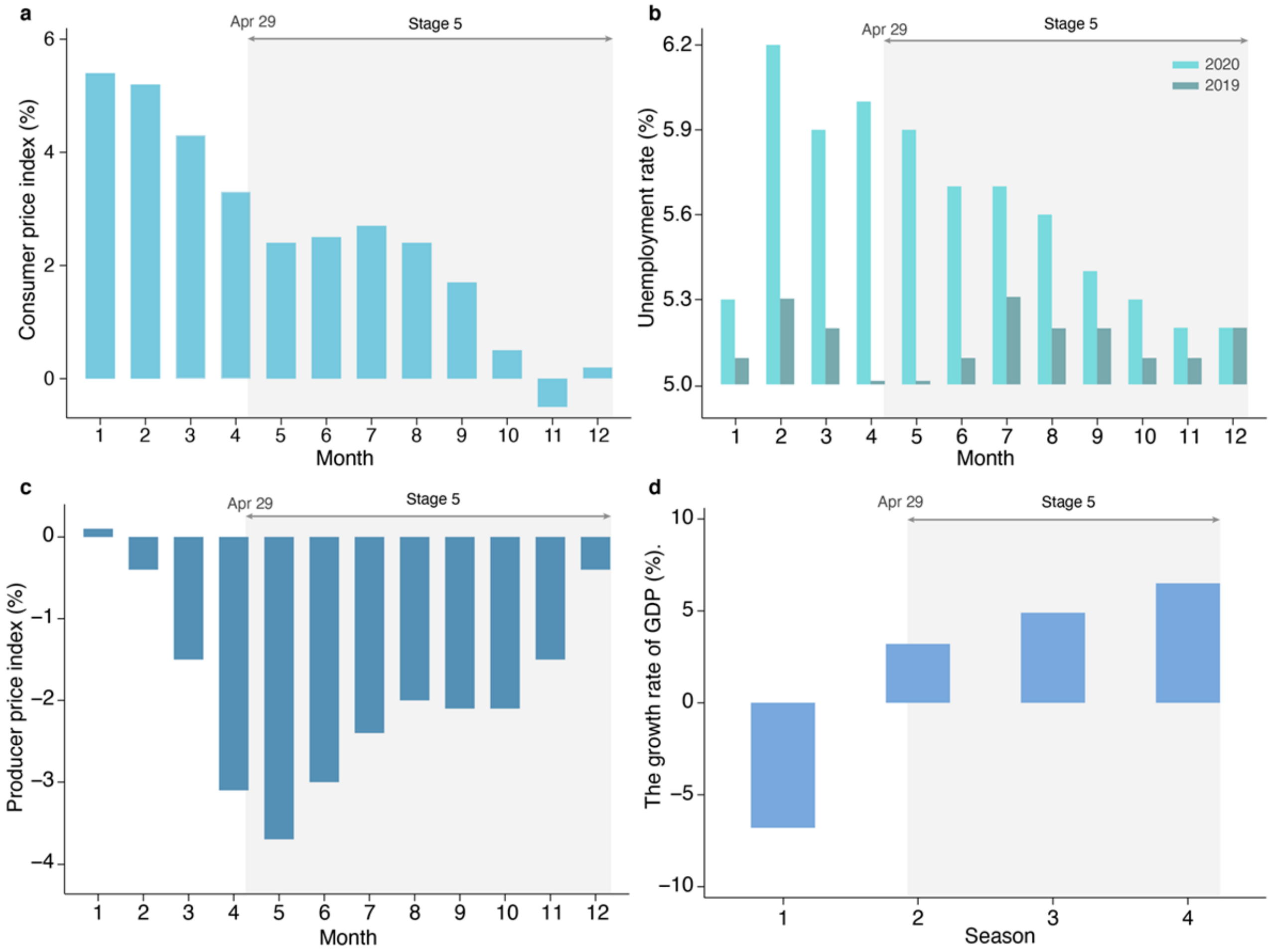
Figure 4. The effect of sporadic indigenous inflection cases in stage 5. (a) National sentiment values in stage 5. The time windows were labeled with underlay colors when the risk assessment level was upgraded to high risk in some areas. (b) Local sentiment values in the provinces when the local risk assessment levels were upgraded.



Concerns About the Economic Consequences of the COVID-19 Pandemic

In stage 5, although the number of newly confirmed COVID-19 cases was relatively lower than earlier in the year, it induced negative effects on economic activities. These negative socioeconomic impacts could lead to public pessimism about economic situations. To test the hypothesis, we computed several social and economic indicators, such as consumer price index, unemployment rate, producer price index, and growth rate of GDP. The consumer price index had fallen since March and was in line with the previous year's level by the end of 2020 (Figure 5a). The unemployment rate also fell to 5.2 at the end of the year, which was lower than the high point of 6.2 at the beginning of the year (Figure 5b). Moreover, the unemployment rate in December was close to that in the previous years. Since February, the producer price index was lower than that in the corresponding period in the last year owing to the impact of the pandemic. However, the gap narrowed after May and reached about 0.4% in December of 2020 (Figure 5c). Meanwhile, the GDP increased, reaching a growth rate of 3.2%, 4.9%, and 6.5% in quarters 2, 3, and 4, respectively (Figure 5d). All major economic indicators showed largely positive economic trends in mainland China in stage 5, leading us to speculate that the decrease in social media public sentiment indices was probably not because of concerns about the impact of economic factors.

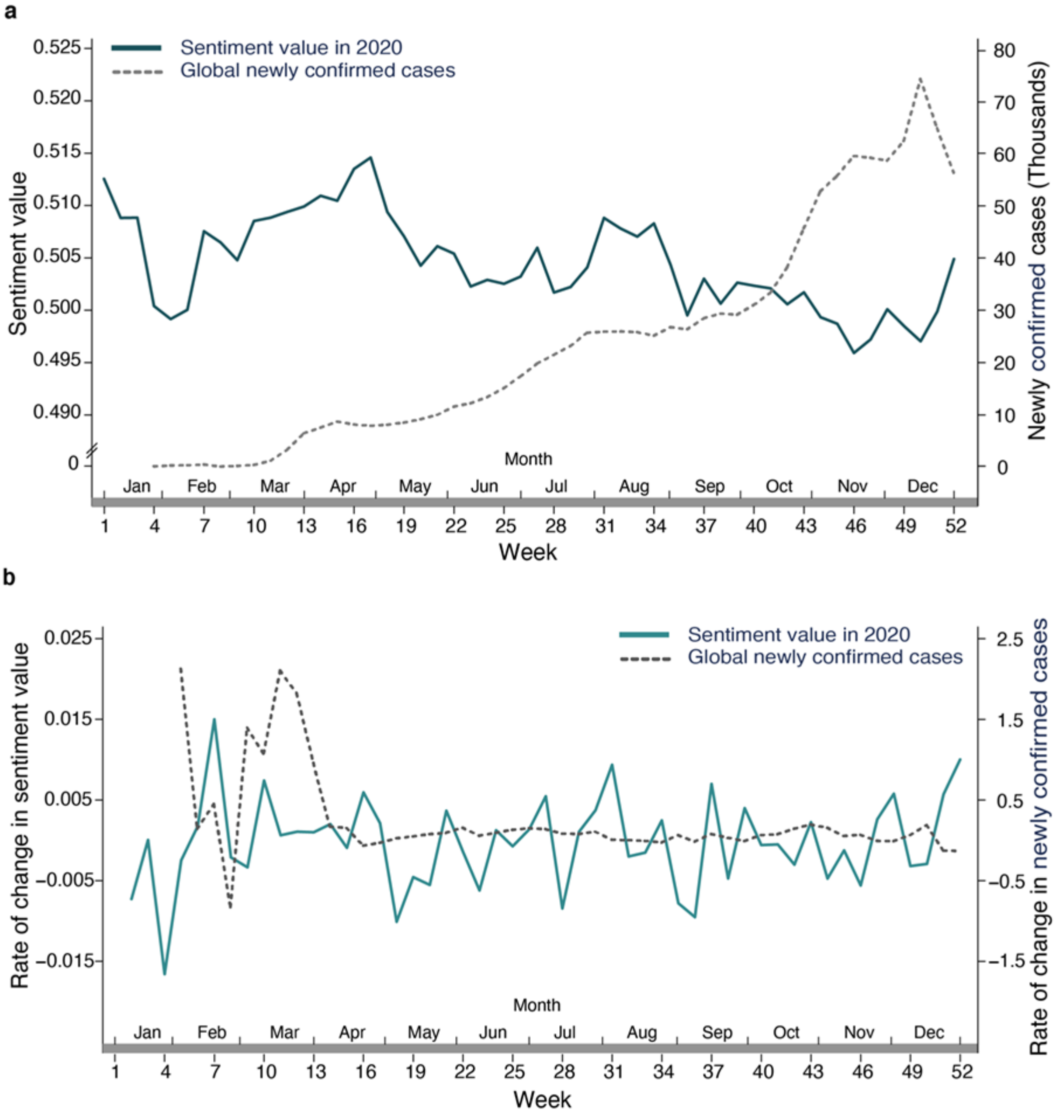
Figure 5. Major economic indicators in 2020 in mainland China, including (a) consumer price index, (b) unemployment rate, (c) producer price index, and (d) gross domestic product (GDP) growth rate.



Concerns About the Severity of the Global Pandemic

In stage 5, although mainland China entered a recovery period, the global pandemic had been worsening. This led to considerations of whether the decline in public sentiment on social media during stage 5 was due to the increasing severity of the global pandemic. To answer this, we compared the sentiment values with the number of newly confirmed COVID-19 cases worldwide in stage 5. Although newly confirmed cases increased while sentiment values decreased in stage 5, timelines of global newly confirmed cases of COVID-19 and sentiment values were out of sync. We calculated the rate of change of newly confirmed cases globally and sentiment values by taking weekly data points. We first subtracted the data point 1 week earlier and then divided the value by the data point 1 week earlier. We observed the rate of change of newly confirmed cases globally and sentiment values did not match (r=0.03, P=.85; see Figure 6b). This analysis indicates that the severity of the global pandemic may not be the reason for the decline in social media public sentiment values in China during stage 5.

Figure 6. Comparisons of (a) global newly confirmed COVID-19 cases with social media public sentiment values and (b) the rates of change between newly confirmed cases globally and sentiment values.



Discussion

Principal Findings

This study tracks public sentiment on social media in mainland China throughout the year of 2020 across the five official stages of the COVID-19 pandemic—from *city lockdown* to *the new normal*. Such uninterrupted long-term tracking allows us to obtain a comprehensive picture of the pandemic's impact on public sentiment on social media. Moreover, we analyzed the data from the corresponding period in the previous two years, while excluding the influence of other possible confounding factors such as holidays. Our analyses identified an interesting new phenomenon—a double descent of public sentiment on social media during the pandemic. The first descent verifies a strong negative effect of public sentiment at the outbreak of COVID-19. More importantly, the second descent illustrates the impact of the pandemic on public sentiment on social media during the recovery stage, which has not been previously reported in the literature.

In terms of the relationship between public sentiment on social media and the severity of the pandemic, there are two separate phases with the boundary at week 17. The first phase, week 1 to week 17, covers stages 1 to 4, defined by the white paper published by the Chinese government. Our study corroborates previous studies, validating that public sentiment was strongly affected by the outbreak of COVID-19 and it was correlated with the severity of the pandemic [33-35]. Moreover, we observed that the public sentiment values increased as China brought the pandemic under control. Especially when the city of Wuhan was reopening, the public sentiment on social media index was even higher than that in the same period in the previous two years, showing a highly positive emotion among the general public after the recovery from the pandemic [16]. These consistencies support that the sentiment analysis of Weibo posts accurately reflects public sentiment during the pandemic.

Public sentiment after week 17, which was defined as stage 5 in the Chinese government's white paper on epidemic control, has not been investigated in previous studies. In this stage, the government had brought the spread of the pandemic under control and socioeconomic life had recovered substantially. We observed a surprisingly sustained decline in social media public sentiment values in stage 5, which differs from the decline in stage 2 in several aspects. First, the public sentiment values in stage 5 were decoupled from the severity of the pandemic. Second, the decline observed in stage 2 lasted only 3 weeks, whereas the decline in stage 5 lasted for at least 34 weeks, with a much slower reduction rate. Third, the social media public sentiment values in stage 5 decreased as a descending spiral rather than a straight line. These characteristics indicate that the decline in stage 5 is unlikely to be a real-time reaction induced by a single event, instead of reflecting a long-term emotional trend in the general public.

Cause Analysis

The underlying causes of such a prolonged decline in public sentiment in stage 5 need to be evaluated. After excluding possible economic reasons and comparing the extent of reduction between public sentiment values in stage 2 and stage 5, we can speculate that the decline could be a long-term effect of the emotional shock of the COVID-19 outbreak. Psychological studies have shown that people exposed to a traumatic event, such as an earthquake, tsunami, or terrorist attack, often struggle with symptoms of posttraumatic stress disorder (PTSD) [36,37]. It is expected that a severe pandemic such as COVID-19 would induce PTSD among many groups, including infected patients [38,39], medical workers [40,41], people related to them, and those who are forced into isolation [42]. However, it is somewhat surprising that there is such a strong and long-last negative effect of the sentiment among the general population. This finding reflects the far-reaching public psychological impact of this unprecedented pandemic.

Limitations and Future Research

Despite the efficiency of using social media for analyzing public sentiment on social media, there remains a gap between the sentiment obtained from texts on the web and real emotion in the general population. Such a gap derives from the sampling bias between the users of the microblogging network and the population in the social group. Although internet access has become widespread in China and Weibo is the largest social media platform there, Weibo users do not represent an unbiased sample of the overall Chinese population because the population of internet users is relatively young and concentrated in metropolitan centers [43]. Moreover, there may also be a difference between the emotions people write in public media and their internal emotions. At the same time, our findings are based on Chinese social media data. At the current stage, mainland China is the only major world economy that has experienced a relatively complete cycle from early outbreaks to the recovery of socioeconomic activities. It is unclear how public sentiment changes in other regions when they enter the recovery phase of the pandemic.

Weibo’s data reflect people's emotions in public scenarios. People may not express their emotions freely to the public due to various concerns. In contrast, the private messages among friends could reflect their internal emotions. In China, there is a popular application named WeChat, which provides text messaging and broadcast messaging among friends. Analysis of WeChat messages may provide public emotions from the other perspective. Moreover, comparing the data of Weibo and WeChat would also show the difference in people's emotional expressions between public and private scenarios. However, WeChat messages are private data and not publicly available, and it would not meet the ethical requirements of using data without the permission of account owners. It would be very valuable to analyze WeChat data if there is a solution to use the data without ethical violations and invasion of personal privacy.

The long-term psychological effects of the COVID-19 pandemic are going to be quite complex and far reaching. Further research will track the effects of the epidemic in the scale of years. Meanwhile, with increasing immunization rates, the epidemic has been sufficiently under control in many other countries. It would be interesting to compare the public emotions across countries with the different cultures after the pandemic. Moreover, given the fact that different countries implement different policies to prevent the spread of the pandemic, it is important to analyze the effects of social policies on public sentiments. These studies would provide very useful information for psychological interventions and social warnings in the post-epidemic world.

Conclusions

In summary, by tracking public sentiment on social media for the whole year of 2020, we were able to evaluate the long-term negative impact of COVID-19 on public sentiment, which shows the complexity and far-reaching impact of the pandemic on human emotions. This study’s results suggest that, from a public policy perspective, even when the pandemic has been controlled and socioeconomic activation is restored, decision-makers must still pay attention to public sentiment and take necessary action to alleviate the negative emotions induced by the pandemic.

Acknowledgments

This research was supported by the Hunan Key Research and Development Project Grant 2020SK2094, the National Key Technologies R&D Program of China Grant 2015BAH22F01 and National Natural Science Foundation of China Grants (No. 31771209).

Conflicts of Interest

None declared.

References

1. WHO Coronavirus Disease (COVID-19) Dashboard. World Health Organization. URL: https://covid19.who.int/ [accessed 2021-03-02]
2. Hawryluck L, Gold WL, Robinson S, Pogorski S, Galea S, Styra R. SARS control and psychological effects of quarantine, Toronto, Canada. Emerg Infect Dis 2004 Jul;10(7):1206-1212 [FREE Full text] [doi: 10.3201/eid1007.030703] [Medline: 15324539]
3. Thompson RR, Garfin DR, Holman EA, Silver RC. Distress, worry, and functioning following a global health crisis: a national study of Americans’ responses to Ebola. Clin Psychol Sci 2017 Apr 26;5(3):513-521. [doi: 10.1177/2167702617692030]
4. Mak IWC, Chu CM, Pan PC, Yiu MGC, Chan VL. Long-term psychiatric morbidities among SARS survivors. Gen Hosp Psychiatry 2009;31(4):318-326 [FREE Full text] [doi: 10.1016/j.genhosppsych.2009.03.001] [Medline: 19555791]
5. Brooks SK, Webster RK, Smith LE, Woodland L, Wessely S, Greenberg N, et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. The Lancet 2020 Mar;395(10227):912-920. [doi: 10.1016/s0140-6736(20)30460-8]
6. Fancourt D, Steptoe A, Bu F. Trajectories of anxiety and depressive symptoms during enforced isolation due to COVID-19 in England: a longitudinal observational study. The Lancet Psychiatry 2021 Feb;8(2):141-149. [doi: 10.1016/s2215-0366(20)30482-x]
7. Holman EA, Thompson RR, Garfin DR, Silver RC. The unfolding COVID-19 pandemic: a probability-based, nationally representative study of mental health in the United States. Sci Adv 2020 Oct 18;6(42):eabd5390 [FREE Full text] [doi: 10.1126/sciadv.abd5390] [Medline: 32948511]
8. Wang Y, Shi L, Que J, Lu Q, Liu L, Lu Z, et al. The impact of quarantine on mental health status among general population in China during the COVID-19 pandemic. Mol Psychiatry 2021 Jan 22:1-10 [FREE Full text] [doi: 10.1038/s41380-021-01019-y] [Medline: 33483692]
9. Pierce M, Hope H, Ford T, Hatch S, Hotopf M, John A, et al. Mental health before and during the COVID-19 pandemic: a longitudinal probability sample survey of the UK population. The Lancet Psychiatry 2020 Oct;7(10):883-892. [doi: 10.1016/s2215-0366(20)30308-4]
10. Li S, Wang Y, Xue J, Zhao N, Zhu T. The impact of COVID-19 epidemic declaration on psychological consequences: a study on active Weibo users. Int J Environ Res Public Health 2020 Mar 19;17(6):2032 [FREE Full text] [doi: 10.3390/ijerph17062032] [Medline: 32204411]
11. Su Y, Wu P, Li S, Xue J, Zhu T. Public emotion responses during COVID‐19 in China on social media: an observational study. Human Behav and Emerg Tech 2020 Dec 10;3(1):127-136. [doi: 10.1002/hbe2.239]
12. Foa R, Gilbert S, Fabian M. COVID-19 and subjective well-being: separating the effects of lockdowns from the pandemic (8/12/2020). SSRN Preprints. Preprint posted online on October 6, 2020 [FREE Full text] [doi: 10.2139/ssrn.3674080]
13. Jacobson NC, Lekkas D, Price G, Heinz MV, 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]
14. Gopal A, Sharma AJ, Subramanyam MA. Dynamics of psychological responses to COVID-19 in India: a longitudinal study. PLoS One 2020 Oct 13;15(10):e0240650 [FREE Full text] [doi: 10.1371/journal.pone.0240650] [Medline: 33048979]
15. Planchuelo-Gómez Á, Odriozola-González P, Irurtia MJ, de Luis-García R. Longitudinal evaluation of the psychological impact of the COVID-19 crisis in Spain. J Affect Disord 2020 Dec 01;277:842-849 [FREE Full text] [doi: 10.1016/j.jad.2020.09.018] [Medline: 33065825]
16. Zhou T, Nguyen TT, Zhong J, Liu J. A COVID-19 descriptive study of life after lockdown in Wuhan, China. R Soc Open Sci 2020 Sep;7(9):200705 [FREE Full text] [doi: 10.1098/rsos.200705] [Medline: 33047032]
17. Lwin MO, Lu J, Sheldenkar A, Schulz PJ, Shin W, Gupta R, et al. Global sentiments surrounding the COVID-19 pandemic on Twitter: analysis of Twitter trends. JMIR Public Health Surveill 2020 May 22;6(2):e19447 [FREE Full text] [doi: 10.2196/19447] [Medline: 32412418]
18. Gupta R, Vishwanath A, Yang Y. Covid-19 Twitter dataset with latent topics, sentiments and emotions attributes. arXiv. Preprint posted online on July 14, 2020 [FREE Full text]
19. Saha K, Torous J, Caine ED, De Choudhury M. Psychosocial effects of the COVID-19 pandemic: large-scale quasi-experimental study on social media. J Med Internet Res 2020 Nov 24;22(11):e22600 [FREE Full text] [doi: 10.2196/22600] [Medline: 33156805]
20. Xue J, Chen J, Chen C, Zheng C, Li S, Zhu T. Public discourse and sentiment during the COVID 19 pandemic: using Latent Dirichlet Allocation for topic modeling on Twitter. PLoS One 2020;15(9):e0239441 [FREE Full text] [doi: 10.1371/journal.pone.0239441] [Medline: 32976519]
21. Wang J, Zhou Y, Zhang W, Evans R, Zhu C. Concerns expressed by chinese social media users during the COVID-19 pandemic: content analysis of Sina Weibo microblogging data. J Med Internet Res 2020 Nov 26;22(11):e22152 [FREE Full text] [doi: 10.2196/22152] [Medline: 33151894]
22. Yu S, Eisenman D, Han Z. Temporal dynamics of public emotions during the COVID-19 pandemic at the epicenter of the outbreak: sentiment analysis of Weibo posts From Wuhan. J Med Internet Res 2021 Mar 18;23(3):e27078 [FREE Full text] [doi: 10.2196/27078] [Medline: 33661755]

23. National Economy Recovered Steadily in 2020 with Main Goals Accomplished Better Than Expectation. Webpage in Chinese. National Bureau of Statistics. 2021 Jan 18. URL: http://www.stats.gov.cn/tjsj/zxfb/202101/t20210118_1812423.html [accessed 2021-03-02]
24. Fighting Covid-19 China in Action. Webpage in Chinese. The State Council Information Office of the People's Republic of China. 2020 Jun. URL: http://www.scio.gov.cn/ztk/dtzt/42313/43142/index.htm [accessed 2021-03-02]
25. Sina Weibo. URL: https://weibo.com/ [accessed 2021-03-02]
26. Li L, Li A, Hao B, Guan Z, Zhu T. Predicting active users' personality based on micro-blogging behaviors. PLoS One 2014;9(1):e84997 [FREE Full text] [doi: 10.1371/journal.pone.0084997] [Medline: 24465462]
27. Natural language processing. Webpage in Chinese. Tencent. URL: https://cloud.tencent.com/product/nlp [accessed 2021-03-02]
28. Li X, Bing L, Li P, Lam W. A unified model for opinion target extraction and target sentiment prediction. In: Proceedings of the AAAI Conference on Artificial Intelligence.: AAAI Press; 2019 Jul 17 Presented at: Thirty-Third AAAI Conference on Artificial Intelligence; 27 January to February 1, 2019; Honolulu, HI p. 6714-6721. [doi: 10.1609/aaai.v33i01.33016714]
29. Epidemic notification. Webpage in Chinese. National Health Commission of the People's Republic of China. URL: http://www.nhc.gov.cn/xcs/yqtb/list_gzbd.shtml [accessed 2021-03-02]
30. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. The Lancet Infectious Diseases 2020 May;20(5):533-534. [doi: 10.1016/s1473-3099(20)30120-1]
31. National data. Webpage in Chinese. National Bureau of Statistics. URL: http://www.stats.gov.cn/english/ [accessed 2021-03-02]
32. Announcement of Wuhan Headquarters for Prevention and Control of Pneumonia Epidemic Caused by Novel Coronavirus Infection (No. 1). Central People's Government of the People's Republic of China. 2020 Jan 23. URL: http://www.gov.cn/xinwen/2020-01/23/content_5471751.htm [accessed 2021-03-02]
33. Burke T, Berry A, Taylor LK, Stafford O, Murphy E, Shevlin M, et al. Increased psychological distress during COVID-19 and quarantine in Ireland: a national survey. J Clin Med 2020 Oct 28;9(11):3481 [FREE Full text] [doi: 10.3390/jcm9113481] [Medline: 33126707]
34. van Agteren J, Bartholomaeus J, Fassnacht DB, Iasiello M, Ali K, Lo L, et al. Using internet-based psychological measurement to capture the deteriorating community mental health profile during COVID-19: observational study. JMIR Ment Health 2020 Jun 11;7(6):e20696 [FREE Full text] [doi: 10.2196/20696] [Medline: 32490845]
35. Kikuchi H, Machida M, Nakamura I, Saito R, Odagiri Y, Kojima T, et al. Changes in psychological distress during the COVID-19 pandemic in Japan: a longitudinal study. J Epidemiol 2020;30(11):522-528. [doi: 10.2188/jea.je20200271]
36. Brewin CR, Holmes EA. Psychological theories of posttraumatic stress disorder. Clin Psychol Rev 2003 May;23(3):339-376. [doi: 10.1016/s0272-7358(03)00033-3]
37. Jordan HT, Osahan S, Li J, Stein CR, Friedman SM, Brackbill RM, et al. Persistent mental and physical health impact of exposure to the September 11, 2001 World Trade Center terrorist attacks. Environ Health 2019 Feb 12;18(1):12. [doi: 10.1186/s12940-019-0449-7] [Medline: 30755198]
38. Kaseda ET, Levine AJ. Post-traumatic stress disorder: a differential diagnostic consideration for COVID-19 survivors. Clin Neuropsychol 2020;34(7-8):1498-1514. [doi: 10.1080/13854046.2020.1811894] [Medline: 32847484]
39. Xiao S, Luo D, Xiao Y. Survivors of COVID-19 are at high risk of posttraumatic stress disorder. Glob Health Res Policy 2020;5:29 [FREE Full text] [doi: 10.1186/s41256-020-00155-2] [Medline: 32514428]
40. Zhang H, Shi Y, Jing P, Zhan P, Fang Y, Wang F. Posttraumatic stress disorder symptoms in healthcare workers after the peak of the COVID-19 outbreak: A survey of a large tertiary care hospital in Wuhan. Psychiatry Res 2020 Dec;294:113541 [FREE Full text] [doi: 10.1016/j.psychres.2020.113541] [Medline: 33128999]
41. Carmassi C, Foghi C, Dell'Oste V, Cordone A, Bertelloni CA, Bui E, et al. PTSD symptoms in healthcare workers facing the three coronavirus outbreaks: what can we expect after the COVID-19 pandemic. Psychiatry Res 2020 Oct;292:113312 [FREE Full text] [doi: 10.1016/j.psychres.2020.113312] [Medline: 32717711]
42. Tang W, Hu T, Hu B, Jin C, Wang G, Xie C, et al. Prevalence and correlates of PTSD and depressive symptoms one month after the outbreak of the COVID-19 epidemic in a sample of home-quarantined Chinese university students. J Affect Disord 2020 Sep 01;274:1-7 [FREE Full text] [doi: 10.1016/j.jad.2020.05.009] [Medline: 32405111]
43. 2018 Weibo User Development Report. Webpage in Chinese. Sina Weibo Data Center. 2019 Mar 15. URL: https://data.weibo.com/report/reportDetail?id=433 [accessed 2021-03-02]

Abbreviations

GDP: gross domestic product
SARS: severe acute respiratory syndrome

Edited by C Basch; submitted 28.03.21; peer-reviewed by S Gordon, Y Yu, S Molani; comments to author 08.06.21; revised version received 10.07.21; accepted 12.07.21; published 12.08.21.

Please cite as:
Tan H, Peng SL, Zhu CP, You Z, Miao MC, Kuai SG
Long-term Effects of the COVID-19 Pandemic on Public Sentiments in Mainland China: Sentiment Analysis of Social Media Posts
J Med Internet Res 2021;23(8):e29150
URL: https://www.jmir.org/2021/8/e29150
doi:10.2196/29150
PMID:34280118

©Hao Tan, Sheng-Lan Peng, Chun-Peng Zhu, Zuo You, Ming-Cheng Miao, Shu-Guang Kuai. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 12.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Predictors of COVID-19 Preventive Perceptions and Behaviors Among Millennials: Two Cross-sectional Survey Studies

Christopher E Beaudoin[1*], PhD; Traci Hong[1*], PhD

College of Communication, Boston University, Boston, MA, United States

[*] all authors contributed equally

Corresponding Author:
Christopher E Beaudoin, PhD
College of Communication
Boston University
640 Commonwealth Ave
Boston, MA, 02215
United States
Phone: 1 617 353 3450
Email: cebeau@bu.edu

Abstract

Background: COVID-19 preventive perceptions and behaviors, especially among US millennials, are an important means by which the pandemic can be slowed and negative health outcomes can be averted.

Objective: This manuscript aims to advance knowledge on COVID-19 preventive perceptions and behaviors and their main predictors, including digital health information–seeking behavior (HISB), political party identification, and COVID-19 testing status.

Methods: Two cross-sectional online surveys of US millennials were conducted from April 10 to 14, 2020 (N=274) (ie, Study 1), and from April 27 to May 7, 2020 (N=1037) (ie, Study 2). In the regression models, dependent variables included preventive behaviors (eg, wearing a face mask and social distancing) as well as four preventive perceptions: severity (ie, a person's conception of the seriousness of COVID-19), susceptibility (ie, a person's conception of the likelihood of being infected with COVID-19), self-efficacy (ie, a person's perception that he or she can wear a face mask and perform social distancing to prevent COVID-19 infection), and response efficacy (ie, a person's perception of whether wearing a face mask and social distancing can prevent COVID-19 infection). Key independent variables included digital HISB for self, digital HISB for another person, political party identification, and COVID-19 testing status.

Results: Millennials reported lower levels of perceived susceptibility than the other three preventive perceptions (ie, severity, self-efficacy, and response efficacy), as well as fairly high levels of preventive behaviors. Unlike HISB for another person, digital HISB for self was positively associated with preventive perceptions and behaviors. In Study 1, respondents with higher levels of digital HISB for self had significantly higher perceptions of severity (β=.22, P<.001), self-efficacy (β=.15, P=.02), and response efficacy (β=.25, P<.001) as well as, at nearing significance, higher perceptions of susceptibility (β=.11, P=.07). In Study 2, respondents with higher levels of digital HISB for self had significantly higher perceptions of severity (β=.25, P<.001), susceptibility (β=.14, P<.001), and preventive behaviors (β=.24, P<.001). Preventive behaviors did not vary significantly according to political party identification, but preventive perceptions did. In Study 1, respondents who identified as being more Republican had significantly lower perceptions of self-efficacy (β=−.14, P=.02) and response efficacy (β=−.13, P=.03) and, at nearing significance, lower perceptions of severity (β=−.10, P=.08) and susceptibility (β=−.12, P=.06). In Study 2, respondents who identified as being more Republican had significantly lower perceptions of severity (β=−.08, P=.009). There were mixed effects of COVID-19 testing status on preventive perceptions, with respondents who had tested positive for COVID-19 having significantly higher perceptions of susceptibility in Study 1 (β=.17, P=.006) and significantly lower perceptions of severity in Study 2 (β=−.012, P<.001).

Conclusions: As the largest and most digitally savvy generation, US millennials saw COVID-19 as a severe threat, but one that they were less susceptible to. For millennials, digital HISB for self, but not for another person, was critical to the development of preventive perceptions and behaviors.

(J Med Internet Res 2021;23(8):e30612) doi:10.2196/30612

KEYWORDS

COVID-19; coronavirus; pandemic; preventive perceptions; preventive behaviors; health information seeking; political party identification; COVID-19 testing

Introduction

Background

COVID-19 is a major global health threat, with 155 million global cases and 3.2 million related deaths as of May 6, 2021, including 32 million cases and 580,012 deaths in the United States [1]. Two pandemic time frames are especially pertinent to this manuscript: April 10 to 14, 2020, and April 27 to May 7, 2020. By April 14, there were 1.8 million confirmed coronavirus cases and 120,548 related deaths globally, including 582,594 confirmed cases and 23,649 deaths in the United States [2]. By May 7, there were 3.72 million confirmed coronavirus cases and 263,489 related deaths, including 1.23 million confirmed cases and 73,431 deaths in the United States.

Before and across these time frames, there were school and workplace closures and stay-at-home orders across the United States, as well as recommendations of preventive measures, such as social distancing and face-mask wearing. In its declaration of COVID-19 as a pandemic in March 2020, the World Health Organization labeled the plethora of coronavirus information as an *infodemic* [3], which stresses the importance of health information. Without effective vaccines prior to 2021, a significant challenge for policy and practical intervention has been encouraging individuals to adopt preventive behaviors (eg, wearing face masks and social distancing) as a means of preventing the virus' continued spread and the further escalation of negative outcomes. To identify what types of individuals are most likely—and least likely—to adopt COVID-19 preventive behaviors, this manuscript investigates preventive perceptions and behaviors and their predictors, including digital health information–seeking behavior (HISB), political party identification, and COVID-19 testing status.

With implications for theory, policy, and practical intervention, this study focuses on COVID-19 preventive perceptions and behaviors of US millennials, who have been an especially important population for COVID-19 preventive efforts [4], given their high levels of social activity and tendency to have no or mild symptoms [5]. According to Dr Deborah Birx, the White House coronavirus response coordinator, millennials are "the core group that will stop this virus" [6]. Millennials make up the largest generational segment in the United States in terms of both the population and workforce [7,8] and are considered *digital natives* [9], being lifelong users of internet and digital media and the most active and experienced generation in terms of new and emerging technologies [10].

Predictors of Preventive Perceptions and Behaviors

Theoretical Basis

Health behavioral theory provides a framework for understanding psychological processes by which people confront health threats such as COVID-19. The Extended Parallel Process Model (EPPM) entails two predictive processes: danger control and fear control [11]. EPPM postulates that health messaging can influence four types of preventive perceptions (ie, severity, susceptibility, self-efficacy, and response efficacy), with these effect processes determining whether individuals enter danger control or fear control. Individuals who develop sufficient levels of these four perceptions are expected to enter danger control, develop protection motivation, undergo an adaptive response, and adopt preventive behaviors. In EPPM, preventive behavior is an "action to prevent, detect, or control illness conditions" (page 76 in Skinner et al [12]). COVID-19 preventive behaviors include wearing a face mask, staying at home, avoiding public or crowded places, avoiding travel, avoiding contact with high-risk individuals, washing or sanitizing hands, and social distancing [13,14]. Perceived severity entails a person's conception of a health threat's seriousness, whereas perceived susceptibility involves a person's conception of the likelihood of undergoing the threat [11]. Self-efficacy entails individuals' perceptions that they can perform a suggested response, whereas response efficacy refers to individuals' perceptions of whether a response will prevent a threat [11].

With a basis in EPPM, this manuscript employs the following COVID-19 outcomes: preventive behaviors (ie, avoiding travel, avoiding gatherings, staying at home, wearing a face mask, washing hands, social distancing, and sheltering in place) and four preventive perceptions (ie, severity, susceptibility, self-efficacy, and response efficacy). The four preventive perceptions align with recent research that has modeled COVID-19 preventive behaviors. In particular, studies have identified fear and threat components, such as severity and susceptibility [15,16], as well as self-efficacy [15-17] and response efficacy [16], as instrumental in predicting preventive behaviors. Research has documented US millennials' perceived risk of COVID-19 and practice of social distancing. There is evidence that millennials are less likely to social distance than prior generations, but have higher risk perceptions than other generations [18]. Other research has considered the predictors of COVID-19 preventive perceptions and behaviors, focusing on either HISB [19] or political party identification [20]. To advance this area of research, we hypothesize that both HISB and political party identification are influential in the development of preventive perceptions and behaviors related to infectious disease. We also consider the effects of COVID-19 testing status.

Health Information–Seeking Behavior

HISB entails how individuals purposively seek out health information from media and other sources [21] and is a means to coping with a health threat and emotions that result from the threat [22]. This manuscript's measurement of HISB is specific to digital access via computer, smartphone, or other electronic means. With the advance of new media technologies, HISB has increased among US adults, with their reliance on the internet for health information rising from 41% in 2005 to 66% in 2014 [23]. Among US millennials, internet use expanded from 83% in 2005 to 90% in 2010, which was higher than for other generations [10]. Particular to internet-using US millennials, 92% in 2017 and 91% in 2019 conducted a digital search for health or medical information *for self* [24,25].

Across the COVID-19 pandemic, the internet has been a powerful conduit for COVID-19 information. In one study of social media posts in China, it was documented that the public was attentive to such information, especially given that the pandemic had closed off some normal interpersonal communication conduits [26]. During the pandemic, US adults have commonly accessed coronavirus news and other related mediated information. In late March 2020, 92% reported having *fairly or very closely* followed news about the coronavirus pandemic [27], whereas in early April 2020, 87% reported that the internet had been important or essential to them during the pandemic [28]. HISB is imperative given its documented effects on health-related knowledge, preventive behaviors, social support, and emotional health [29-32]. Specific to US millennials, use of the internet for health information neared saturation (ie, 99.3%) prior to the pandemic, which was markedly higher than for members of Generation X and baby boomers [33]. In the context of COVID-19, one study documented that information receptivity was positively associated with COVID-19 preventive behaviors [15]. Another study found that digital COVID-19 information seeking had positive direct effects on preventive behaviors, as well as indirect effects as mediated by perceived worry [19]. Given this basis in the literature, Hypothesis 1 (H1) states the following: Digital HISB is positively associated with COVID-19 preventive perceptions and behaviors.

The surveys used in this manuscript also permit a nuanced examination of HISB. While most research has operationalized HISB in terms of a person's HISB *for self*, some research has considered a person's HISB *for another person* [34]. HISB for self entails a person seeking out health information to address a personal health risk or threat, whereas HISB for another person entails a person seeking out health information to address another person's health risk or threat. This second type of HISB could entail a person seeking out health information for a family member, friend, or other care recipient. This leads to Research Question 1 (RQ1): Does the relationship between digital HISB and COVID-19 preventive perceptions and behaviors vary according to HISB typology—for self versus for another person?

Political Party Identification

Political party identification entails individuals' identification as Democrat, Republican, or Independent, as well as their strength of identification in the first two regards and their leaning toward Democrat or Republican in the third regard [35]. The opposition of Republicans and Democrats to one another, as well as the divide in related sentiments and mutual antipathy, has broadened in recent decades [36]. This growing division has been instrumental in the politicization of the COVID-19 pandemic, which has manifested itself in public health practices and related misinformation [37] and in the polarization of coverage in partisan news media [38]. Misinformation in the media has led to political divides in terms of what is said by opinion leaders and understood by their constituents, including in regard to whether the pandemic actually exists, whether people should adopt preventive behaviors, whether vaccines are safe, and whether people should get them. In this manner, news and other media content have augmented partisan differences in COVID-19 perceptions and practices.

Exemplifying the partisan nature of COVID-19, research in the United States has found that Republican counties have been much less likely to practice social distancing than Democrat counties, with these county-level differences associated with individuals' viewership of the conservative Fox News channel [20]. Another study documented that, as compared to liberals, conservatives have reported lower levels of fear of COVID-19 [39]. Given this basis, Hypothesis 2 (H2) is as follows: Political party identification is inversely associated with COVID-19 preventive perceptions and behaviors, with respondents who are more Republican exhibiting lower levels than respondents who are more Democrat.

COVID-19 Testing Status

An individual's COVID-19 testing status can be viewed as an indicator of personal relevance [21]. The presence of personal experiences enhances the social significance of related health information and disease processes and can be operationalized in terms of a person's testing and diagnosis for a disease and experiencing of symptomology [21].

Across the pandemic, the implementation of COVID-19 testing has been important to documenting who has the virus. As of May 6, 2021, there had been 439 million tests conducted in the United States [1]. As a predictor of COVID-19 preventive perceptions and behaviors, it is unclear whether coronavirus testing status would have a positive or negative effect. On one hand, it could be that individuals who tested positive would have a higher perception of the relevance of the coronavirus and, thus, may have elevated levels of COVID-19 preventive perceptions and behaviors. On the other hand, it could be that individuals who tested positive and did not experience negative results would have diminished levels of COVID-19 preventive perceptions and behaviors. Given this uncertainty, Research Question 2 (RQ2) is as follows: What is the relationship between testing positive for COVID-19 and COVID-19 preventive perceptions and behaviors?

Methods

Overview

To address the two hypotheses and two research questions, we implemented two empirical studies. For both studies, data were collected via online self-report survey questionnaires. Study 1 examined the effects of digital HISB for self, political party identification, and COVID-19 testing status on four preventive perceptions. Study 2 examined the effects of digital HISB for self, digital HISB for another person, political party identification, and COVID-19 testing status on preventive behaviors and two preventive perceptions. This two-study approach permitted several benefits, including the replication of results specific to the effects of the key antecedents on perceived severity and susceptibility across two separate time frames in April and May 2020. It also permitted the expansion of results to include effects on perceived self-efficacy and response efficacy in Study 1 and to include effects on preventive behaviors in Study 2.

Institutional Review Board approval for both cross-sectional survey studies was attained at the research university of the authors (Study 1: 5550X; Study 2: 5572X). Prior to the online surveys, respondents were provided with an informed consent statement, which included specification of the survey's purpose, length, and investigators and an indication that participation was voluntary and respondents could stop answering questions at any point across the questionnaire. Qualtrics hosted both surveys and recruited and compensated respondents. Qualtrics recruits samples from traditional research panels. To guarantee the validity, reliability, and integrity of survey data, Qualtrics checks all IP addresses and implements digital fingerprinting technology. Qualtrics and its sample partners implement various procedures to confirm respondent identity and randomly select respondents from survey panels who appear to meet a study's articulated population parameters. Crafted in a general manner as a means of decreasing potential selection bias, respondent invitations included email, in-app, and SMS notifications. In these studies, panelists who met the sought-after sampling frame were provided with an opportunity to partake in the survey. After the completion of the survey, Qualtrics created a variable in the data set that it recommended for usage that excluded respondents who refused consent, did not complete the survey, completed it in shorter than one-half the median survey completion time, or did not match the sought-after sample quotas.

Study 1

Sample

The first study used data from a cross-sectional online survey of US adults aged 18 years and older. Interviews were conducted from April 10 to 14, 2020 (N=1014). Qualtrics hosted the survey and derived the sample, aiming to match quotas for the US adult population in terms of age, gender, education, household income, and ethnicity. After excluding incompletes and after Qualtrics adjustments to ensure data quality and to match the sought-after sample quotas (n=393), the final sample was comprised of 1014 completed survey interviews. Given this manuscript's focus on millennials, the subset of respondents aged 25 to 39 years (N=274) was used for all statistical analyses.

Measurement

The survey questionnaire's item wording for key independent and dependent variables is depicted in Textbox 1 [13,14,25,35,36,40-42]. Instructions across the questionnaire referred to the COVID-19 coronavirus and used the term *the coronavirus* in the specific questions. Three survey items for perceived severity (α=.87) and three survey items for perceived susceptibility (α=.83) were adapted from prior research [40]. Responses were rated on a 5-point Likert scale as follows: strongly disagree, 1; disagree, 2; neither agree nor disagree, 3; agree, 4; and strongly agree, 5. Six survey items for perceived self-efficacy (α=.88) and six survey items for perceived response efficacy (α=.89) were adapted from prior research [40]. These items were split across two preventive behaviors: wearing a face mask and social distancing. Responses were rated on a 5-point Likert scale as follows: strongly disagree, 1; disagree, 2; neither agree nor disagree, 3; agree, 4; and strongly agree, 5.

Digital HISB for self was measured specific to the novel coronavirus. Using the basic structure of the HISB question from Health Information National Trends Survey (HINTS) 5, Cycle 3 [25], there was one item that entailed seeking of coronavirus information for oneself *in the past month* via *computer, smartphone, or other electronic means*. Responses were rated on a 6-point scale with six possible responses—never, once, several times, once per week, several times per week, and every day—and, with reference to prior research [43], they were then recoded in terms of days per month (0 to 30).

Other key independent variables included political party identification and COVID-19 testing status. For political party identification [35,36], there was a 6-point scale ranging from 1 (strong Democrat) to 6 (strong Republican). In terms of COVID-19 testing status, a dichotomous (ie, yes or no) measure was instituted for whether respondents had tested positive for COVID-19 [41].

Control demographics included age, education, household income, ethnicity, employment, and gender. In terms of the millennial subsample, age responses were from 25 to 39 years. Education was measured at the ordinal level with seven responses ranging from *less than 8 years* to *postgraduate*. Household income was measured at the ordinal level with nine responses ranging from *US $9999 or less* to *US $200,000 or more*. For ethnicity, a dichotomous variable of *White* or *non-White* was created. For employment, a dichotomous variable for full-time employment was created.

Textbox 1. Survey item wording for key independent and dependent variables.

Key independent variables

COVID-19 testing status (responses were 1 [yes] or 0 [no] for the first question; for the second question, responses were 1 [positive] or 0 [negative]) [41]:

- Have you personally been tested for coronavirus, or not?
- Was your coronavirus test positive or negative?

Digital health information–seeking behavior (HISB) for self (responses were rated on a 6-point scale as follows: never, once, several times, once per week, several times per week, and every day) [25]:

- In the past month, how frequently have you used a computer, smartphone, or other electronic means to look for information about coronavirus FOR YOURSELF?

Digital HISB for another person (responses were rated on a 6-point scale as follows: never, once, several times, once per week, several times per week, and every day) [25,42]:

- In the past month, how frequently have you used a computer, smartphone, or other electronic means to look for information about coronavirus FOR ANOTHER PERSON?

Political party identification (for the first introductory question, responses were Republican, Democrat, Independent, no preference, or other party; for the second question, responses were strong Republican or not very strong Republican; for the third question, responses were strong Democrat or not very strong Democrat; and responses for the fourth question were closer to Republican or closer to Democratic) [35,36]:

- Generally speaking, do you usually think of yourself as a Republican, a Democrat, an Independent, or what?
- (If answered Republican to introductory question) Would you call yourself a strong Republican or a not very strong Republican?
- (If answered Democrat to introductory question) Would you call yourself a strong Democrat or a not very strong Democrat?
- (If answered Independent, no preference, or other party to introductory question) Do you think of yourself as closer to the Republican or Democratic party?

Dependent variables

Severity (responses were rated on a 5-point Likert scale ranging from 1 [strongly disagree] to 5 [strongly agree]) [40]:

- I believe that getting the coronavirus is severe.
- I believe that getting the coronavirus has severe negative consequences.
- I believe that getting the coronavirus is extremely harmful.

Susceptibility (responses were rated on a 5-point Likert scale ranging from 1 [strongly disagree] to 5 [strongly agree]) [40]:

- I am at risk for getting the coronavirus.
- It is likely that I will get the coronavirus.
- It is possible that I will get the coronavirus.

Self-efficacy (responses were rated on a 5-point Likert scale ranging from 1 [strongly disagree] to 5 [strongly agree]) [40]:

- I have the ability to do social distancing to prevent getting the coronavirus.
- I am able to do social distancing to prevent getting the coronavirus.
- I can easily do social distancing to prevent getting the coronavirus.
- I can easily wear a face mask to prevent getting the coronavirus.
- I have the ability to wear a face mask to prevent getting the coronavirus.
- I am able to wear a face mask to prevent getting the coronavirus.

Response efficacy (responses were rated on a 5-point Likert scale ranging from 1 [strongly disagree] to 5 [strongly agree]) [40]:

- Social distancing is effective to prevent getting the coronavirus.
- If I do social distancing, I am less likely to get the coronavirus.
- Social distancing works to prevent getting the coronavirus.
- Wearing a face mask is effective to prevent getting the coronavirus.
- If I wear a face mask, I am less likely to get the coronavirus.

- Wearing a face mask works to prevent getting the coronavirus.

Preventive behaviors (responses were 1 [yes] or 0 [no]) [13,14]:

- Because of the coronavirus outbreak, have you decided NOT to travel or have you changed your travel plans?
- Because of the coronavirus outbreak, have you stayed at home instead of going to work, school, or other regular activities?
- Because of the coronavirus outbreak, have you bought or worn a protective face mask?
- Because of the coronavirus outbreak, have you frequently washed your hands with soap and water?
- Because of the coronavirus outbreak, have you tried to stay at least six feet way from other people when outside of your household?
- Because of the coronavirus outbreak, have you sheltered in place?

Statistical Analysis

Statistical analysis was conducted with Stata 16 (StataCorp LLC). The internal consistency of composite measures was assessed with Cronbach α. Descriptive statistics were calculated and reported for each study variable. To test the predictors of preventive perceptions, we used ordinary least squares (OLS) regression. The dependent variables in the models were severity, susceptibility, self-efficacy, and response efficacy. Independent variables were entered in two hierarchical steps: (1) control demographics and (2) digital HISB for self, political party identification, and tested positive for COVID-19. This second block of variables permitted testing of the hypotheses and answering of the research questions. We calculated variance inflation factors (VIFs) for the regression models to gauge for multicollinearity [44]. For the regression models in Study 1 and Study 2, VIF levels were low (ie, <2.00), which indicates no evidence of multicollinearity [44].

Study 2

Sample

The second study used data from a cross-sectional online survey that targeted millennials, in particular US adults aged 25 to 39 years (N=1037). Interviews were conducted from April 27 to May 7, 2020. Qualtrics hosted the survey and derived the sample, aiming to match quotas for the US adult population in terms of gender, education, and ethnicity. After excluding incompletes and after Qualtrics adjustments to ensure data quality and to match the sought-after sample quotas (n=80), the final sample included 1037 completed survey interviews.

Measurement

The survey questionnaire’s item wording for key independent and dependent variables is depicted in Textbox 1. Consistent with Study 1, three survey items for perceived severity (α=.84) and three survey items for perceived susceptibility (α=.76) were adapted from prior research [40] and had the same Likert scale response options as described above. Preventive behaviors were measured with a 7-item additive index (α=.79), with dichotomous (ie, yes or no) questions specific to the following practices: avoiding travel, avoiding gatherings, staying at home, wearing face masks, washing hands, social distancing, and sheltering in place [13,14].

Digital HISB was measured in two ways: for self and for another person. Using the basic structure of the question on HISB for self from HINTS 5, Cycle 3 [25], the measurement was expanded to include HISB for another person, with adaption from HINTS 4, Cycle 1 [42]. Like Study 1, the first item specific to *digital HISB for self* entailed seeking of coronavirus information for oneself *in the past month* via *computer, smartphone, or other electronic means*. The second item specific to *digital HISB for another person* entailed seeking of coronavirus information for another person *in the past month* via *computer, smartphone, or other electronic means*. Like Study 1, responses were rated on a 6-point scale and then recoded in terms of days per month (0 to 30) [43].

Political party identification and COVID-19 testing status were measured with the same survey items as in Study 1. Also like Study 1, control demographics included age, education, household income, ethnicity, employment, and gender.

Statistical Analysis

Statistical analysis was similar to Study 1, also using Stata 16 (StataCorp LLC). The internal consistency of composite measures was assessed with Cronbach α, and descriptive statistics were reported for all measures. To test the predictors of preventive perceptions and behaviors, we used OLS regression. Dependent variables in the models were severity, susceptibility, and preventive behaviors. Independent variables were entered in two hierarchical steps: (1) control demographics and (2) digital HISB for self, digital HISB for another person, political party identification, and tested positive for COVID-19. The second block of variables allowed for testing of the hypotheses and answering of the research questions. As with Study 1, there was no evidence of multicollinearity in the regression models, which had low VIF levels (ie, <2.00) [44].

Results

Study 1

Descriptive statistics appear in Table 1. Of the sample, more than 64% were White, and more than 45% were male. The mean age was more than 32 years, the mode for household income was *$50,000 to $74,999*, and the mode for education was *12 years or completed high school*. Finally, more than 51% of the sample reported full-time employment. Among the key predictors, the average for digital HISB for self was 15 days per month. Average levels of the four preventive perceptions were between 3.13 and 4.05 on the 5-point Likert scale.

Table 1. Study 1 descriptive statistics.

Variables	Value (N=274)
Control independent variables	
Age (years), mean (SD)	32.40 (4.29)
Male, n (%)	125 (45.6)
White, n (%)	177 (64.6)
Full-time employment, n (%)	140 (51.1)
Household income (US $), n (%)	
0 to 9999	34 (12.4)
10,000 to 14,999	12 (4.4)
15,000 to 19,999	20 (7.3)
20,000 to 34,999	43 (15.7)
35,000 to 49,999	35 (12.8)
50,000 to 74,999	70 (25.5)
75,000 to 99,999	31 (11.3)
100,000 to 199,999	26 (9.5)
200,000 or more	3 (1.1)
Education, n (%)	
Less than 8 years	4 (1.5)
8 to 11 years	24 (8.8)
12 years or completed high school	120 (43.8)
Post–high school training other than college	21 (7.7)
Some college	45 (16.4)
College degree	42 (15.3)
Postgraduate	18 (6.6)
Key independent variables	
Digital health information–seeking behavior (HISB) for self[a], mean (SD)	15.35 (12.47)
Political party identification[b], mean (SD)	3.18 (1.68)
Tested positive for COVID-19, n (%)	4 (1.5)
Dependent variables[c], mean (SD)	
Severity	4.04 (0.98)
Susceptibility	3.13 (1.02)
Self-efficacy	4.05 (0.92)
Response efficacy	3.98 (0.89)

[a] HISB was rated on a 6-point scale as follows: never, once, several times, once per week, several times per week, and every day. Responses were then recoded in terms of days per month (0 to 30).

[b] Political party identification was rated on a 6-point scale ranging from 1 (strong Democrat) to 6 (strong Republican).

[c] Responses were rated on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree).

Table 2 depicts the hierarchical regression models for this study. The effects of only control variables can be found in Block 1 in Models 1, 3, 5, and 7. At the level of nearing significance (ie, *P*<.10), age was positively associated with severity (see Model 1), White respondents had higher levels of perceived susceptibility (see Model 3), and household income was positively associated with response efficacy (see Model 7). Finally, full-time employed respondents had lower levels of severity, self-efficacy, and response efficacy, with the final relationship nearing significance (see Models 1, 5, and 7).

The regression results in Table 2 also pertain to H1 and H2 as well as to RQ2 (see Block 2 in Models 2, 4, 6, and 8). Supportive of H1, digital HISB for self was positively associated with severity (β=.22, *P*<.001), self-efficacy (β=.15, *P*=.02), and response efficacy (β=.25, *P*<.001) (see Models 2, 6, and 8).

Also, there was a near-significant positive association between digital HISB for self and susceptibility (β=.11, *P*=.07) (see Model 4). In each case, respondents with higher levels of digital HISB for self had higher perception levels. Supportive of H2, political party identification was negatively associated with self-efficacy (β=−.14, *P*=.02) and response efficacy (β=−.13, *P*=.03) (see Models 6 and 8). Notably, political party identification also had near-significant inverse associations with severity (β=−.10, *P*=.08) and susceptibility (β=−.12, *P*=.06) (see Models 2 and 4). In each of the four cases, respondents who identified as being more Republican had lower perception levels. Finally, in terms of RQ2, respondents who tested positive for COVID-19 had higher levels of susceptibility (β=.17, *P*=.006) (see Model 4).

Table 2. Hierarchical regression predictors of COVID-19 preventive perceptions (N=274).

Independent variable	Dependent variable															
	Severity[a]				Susceptibility[b]				Self-efficacy[c]				Response efficacy[d]			
	M[e]1	*P* value	M2	*P* value	M3	*P* value	M4	*P* value	M5	*P* value	M6	*P* value	M7	*P* value	M8	*P* value
Block 1																
Male, β[f]	–.05	.47	–.04	.56	.02	.76	.02	.72	–.01	.93	.00	.97	–.01	.87	.00	.99
White, β	.05	.42	.09	.13	.12	.06	.15	.01	.05	.39	.10	.12	.05	.38	.10	.09
Household income, β	–.01	.91	–.03	.67	.05	.48	.05	.55	.05	.51	.04	.58	.14	.07	.12	.13
Education, β	.07	.32	.04	.62	.08	.29	.05	.49	.07	.33	.04	.60	–.04	.56	–.09	.23
Age, β	.11	.07	.09	.13	.02	.70	–.01	.86	.07	.24	.05	.46	.06	.34	.03	.58
Full-time employment, β	–.18	.10	–.15	.02	.07	.34	.06	.38	–.18	.01	–.17	.01	–.13	.06	–.11	.12
R^2	0.05	N/A[g]	N/A	N/A	0.04	N/A	N/A	N/A	0.04	N/A	N/A	N/A	0.03	N/A	N/A	N/A
F	2.48	<.001	N/A	N/A	1.76	.11	N/A	N/A	1.86	.09	N/A	N/A	1.52	.17	N/A	N/A
Block 2																
Digital health information–seeking behavior for self, β	N/A	N/A	.22	<.001	N/A	N/A	.11	.07	N/A	N/A	.15	.02	N/A	N/A	.25	<.001
Political party identification, β	N/A	N/A	–.10	.08	N/A	N/A	–.12	.06	N/A	N/A	–.14	.02	N/A	N/A	–.13	.03
Tested positive for COVID-19, β	N/A	N/A	.05	.42	N/A	N/A	.17	.006	N/A	N/A	.09	.15	N/A	N/A	.07	.23
R^2	N/A	N/A	0.11	N/A	N/A	N/A	0.09	N/A	N/A	N/A	0.09	N/A	N/A	N/A	0.11	N/A
ΔR^2	N/A	N/A	0.06	N/A	N/A	N/A	0.05	N/A	N/A	N/A	4.58	N/A	N/A	N/A	0.08	N/A
ΔF	N/A	N/A	5.91	<.001	N/A	N/A	5.28	<.001	N/A	N/A	4.19	.01	N/A	N/A	8.12	<.001

[a] With severity as the dependent variable, the two hierarchical models are Models 1 and 2.

[b] With susceptibility as the dependent variable, the two hierarchical models are Models 3 and 4.

[c] With self-efficacy as the dependent variable, the two hierarchical models are Models 5 and 6.

[d] With response efficacy as the dependent variable, the two hierarchical models are Models 7 and 8.

[e] M: Model.

[f] β values represent the standardized coefficients.

[g] N/A: not applicable for the first model in each dependent variable’s pair of hierarchical regression models, or not applicable for reported R^2, ΔR^2, F, or ΔF.

Study 2

Descriptive statistics appear in Table 3. Of the sample, more than 63% were White and more than 46% were male. The mean age was almost 32 years, the mode for household income was *US $20,000 to US $34,999*, and the mode for education was *12 years or completed high school*. Finally, 52.3% of the sample reported full-time employment. Among the key predictors, the mean for digital HISB for self was more than 13 days per month, whereas the mean for digital HISB for another person was more than 8 days per month. Average levels of severity and self-efficacy were between 3 and 4 on the 5-point Likert scale, and respondents reported having performed an average of almost six of the seven preventive behaviors, with frequently washing hands ranking the highest.

Table 4 depicts the hierarchical regression models for this study. The effects of only control variables can be found in Block 1 in Models 1, 3, and 5. Household income had significant positive associations with susceptibility and preventive behaviors (see Models 3 and 5). At the level of nearing significance, education and age had positive associations with severity (see Model 1). In addition, education was positively associated with preventive behaviors, and, nearing significance, employment was inversely associated with preventive behaviors (see Model 5).

The regression results in Table 4 also pertain to H1 and H2 as well as to RQ1 and RQ2 (see Block 2 in Models 2, 4, and 6). The results for digital HISB for self and digital HISB for another person relate to H1. Support is limited to digital HISB for self, which was positively associated with severity (β=.25, *P*<.001), susceptibility (β=.14, *P*<.001), and preventive behaviors (β=.24, *P*<.001) (see Models 2, 4, and 6). The effects of digital HISB for another person were nonsignificant in each of these models. Thus, in terms of RQ1, digital HISB for self had significant positive associations with each outcome, while digital HISB for another person did not. Supportive of H2, political party identification had a negative association with severity (β=−.08, *P*=.009) (see Model 2). Thus, respondents who identified as being more Republican had lower perceptions of severity. Political party identification did not have significant associations with perceived susceptibility or preventive behaviors. In terms of RQ2, respondents who had tested positive for COVID-19 reported lower levels of perceived severity (β=−.012, *P*<.001) (see Model 2).

Table 3. Study 2 descriptive statistics.

Variables	Value (N=1037)
Control independent variables	
Age (years), mean (SD)	31.65 (4.18)
Male, n (%)	483 (46.58)
White, n (%)	657 (63.36)
Full-time employment, n (%)	542 (52.27)
Household income (US $), n (%)	
0 to 9999	142 (13.69)
10,000 to 14,999	59 (5.69)
15,000 to 19,999	65 (6.27)
20,000 to 34,999	191 (18.42)
35,000 to 49,999	171 (16.49)
50,000 to 74,999	171 (16.49)
75,000 to 99,999	118 (11.38)
100,000 to 199,999	86 (8.29)
200,000 or more	34 (3.28)
Education, n (%)	
Less than 8 years	16 (1.54)
8 to 11 years	52 (6.56)
12 years or completed high school	358 (34.52)
Post–high school training other than college	98 (9.45)
Some college	198 (19.09)
College degree	246 (23.72)
Postgraduate	69 (6.65)
Key independent variables	
Digital health information–seeking behavior (HISB) for self[a], mean (SD)	13.14 (12.28)
HISB for another person[a], mean (SD)	8.14 (10.33)
Political party identification[b], mean (SD)	3.23 (1.68)
Tested positive for COVID-19, n (%)	58 (5.59)
Dependent variables	
Severity[c], mean (SD)	3.84 (1.07)
Susceptibility[c], mean (SD)	3.09 (1.03)
Preventive behaviors	
7-item index[d], mean (SD)	5.86 (1.70)
Avoided travel, n (%)	831 (80.14)
Canceled large gatherings, n (%)	857 (82.64)
Stayed at home, n (%)	830 (80.04)
Wore a protective face mask, n (%)	841 (81.10)
Frequently washed hands, n (%)	941 (90.74)
Stayed six feet away from others, n (%)	911 (87.85)
Sheltered in place, n (%)	850 (81.97)

[a] HISB was rated on a 6-point scale as follows: never, once, several times, once per week, several times per week, and every day. Responses were then

recoded in terms of days per month (0 to 30).

[b] Political party identification was rated on a 6-point scale ranging from 1 (strong Democrat) to 6 (strong Republican).

[c] Responses were rated on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree).

[d] The 7-item index is an additive composite variable of the seven individual preventive behaviors.

Table 4. Hierarchical regression predictors of COVID-19 preventive perceptions and behaviors (N=1037).

Independent variable	Dependent variable											
	Severity[a]				Susceptibility[b]				Preventive behaviors[c]			
	M[d]1	*P* value	M2	*P* value	M3	*P* value	M4	*P* value	M5	*P* value	M6	*P* value
Block 1												
Male, β[e]	–.05	.15	–.05	.12	.00	.89	–.01	.86	–.02	.51	–.03	.36
White, β	.04	.22	.05	.10	.04	.23	.05	.11	–.03	.34	–.02	.46
Household income, β	.05	.15	.05	.11	.09	.01	.09	.01	.08	.02	.09	.01
Education, β	.06	.06	.01	.68	.04	.32	.00	.90	.17	<.001	.14	<.001
Age, β	.06	.06	.04	.22	.01	.75	.01	.81	.01	.63	.00	.88
Full-time employment, β	–.05	.17	–.04	.22	–.01	.78	–.02	.56	–.06	.08	–.06	.05
R^2	0.02	N/A[f]	N/A	N/A	0.01	N/A	N/A	N/A	0.05	N/A	N/A	N/A
F	2.68	.01	N/A	N/A	2.06	.06	N/A	N/A	8.59	<.001	N/A	N/A
Block 2												
Digital health information–seeking behavior (HISB) for self, β	N/A	N/A	.25	<.001	N/A	N/A	.14	<.001	N/A	N/A	.24	<.001
Digital HISB for another person, β	N/A	N/A	.04	.24	N/A	N/A	.06	.10	N/A	N/A	.00	.95
Political party identification, β	N/A	N/A	–.08	.009	N/A	N/A	–.06	.08	N/A	N/A	–.03	.28
Tested positive for COVID-19, β	N/A	N/A	–.12	<.001	N/A	N/A	.06	.06	N/A	N/A	.02	.57
R^2	N/A	N/A	0.12	N/A	N/A	N/A	0.05	N/A	N/A	N/A	0.11	N/A
ΔR^2	N/A	N/A	0.10	N/A	N/A	N/A	0.04	N/A	N/A	N/A	0.06	N/A
ΔF	N/A	N/A	29.45	<.001	N/A	N/A	1.19	<.001	N/A	N/A	17.18	<.001

[a] With severity as the dependent variable, the two hierarchical models are Models 1 and 2.

[b] With susceptibility as the dependent variable, the two hierarchical models are Models 3 and 4.

[c] With preventive behaviors as the dependent variable, the two hierarchical models are Models 5 and 6.

[d] M: Model.

[e] β values represent the standardized coefficients.

[f] N/A: not applicable for the first model in each dependent variable's pair of hierarchical regression models, or not applicable for reported R^2, ΔR^2, F, or ΔF.

Discussion

Principal Findings

Our analyses depict US millennials' levels of COVID-19 preventive perceptions and behaviors. Danger control, the beneficial process entailing protection motivation and behavior change, requires heightened levels of preventive perceptions: severity, susceptibility, self-efficacy, and response efficacy [11]. Of these four perceptions, levels of susceptibility were lowest in both studies, suggesting a gap in millennials' understanding of their potential vulnerability in contracting COVID-19. It is somewhat surprising that respondents had what seem to be low perceptions of the likelihood that they could be infected with COVID-19 given that, across this study's two time frames (ie, April 10 to 14, 2020, and April 27 to May 7, 2020), the virus was widespread and expected to spread further. Thus, whereas levels of perceived severity, self-efficacy, and response efficacy were near 4.00 on the respective 5-point scales, perceptions of susceptibility lagged behind, posing a gap that could be targeted with preventive messaging. Study 2 also provided a picture of whether millennials had performed COVID-19 preventive behaviors. In particular, more than 80% of millennials reported having performed the seven separate preventive behaviors.

In addition, our analyses identified different types of US millennials who were most likely—as well as least likely—to adopt preventive perceptions and behaviors in the context of the COVID-19 pandemic; in a nutshell, health information seeking matters, at least when it comes to self. Across the two empirical studies, digital HISB for self was positively associated with preventive behaviors and each preventive perception; the association was nearing significance in terms of one of the preventive perceptions (ie, susceptibility) in Study 1, which implemented the smaller of the two survey samples. The positive effects of digital HISB are generally consistent with prior research specific to the COVID-19 pandemic [15,19] and other health contexts [29-32]. For example, research has documented that individuals with higher information receptivity [15] and COVID-19 information seeking [19] are more likely to practice COVID-19 preventive behaviors. The results of our two studies underscore the importance of digital health information in the contemporary media world, where mediated health information is widespread and the technical affordances of digital and social media allow users to search out health information in a manner that is purposive or incidental, unbounded by constraints of location and privacy, and diverse in information content and information sources. These findings are also instructive for practical interventions, suggesting the importance of disseminating credible and purposive preventive information across the pandemic. While there is evidence in the literature of misinformation and its negative influence on COVID-19 knowledge and preventive behaviors [45], the overall effect, as documented in this manuscript, is a favorable one in terms of digital HISB for self. Notably, digital HISB for another person was not a significant predictor of any of the outcomes, which suggests that millennials' danger control processes are a function of their drive to protect themselves, not other people [11]. That the related processes of millennials are driven by digital HISB for self, but not digital HISB for another person, may relate to research that has documented millennials as being more self-centered than previous generations [46].

Political party identification had negative coefficients in each regression model for preventive perceptions, achieving significance in three models and nearing significance in the other three. In each case, respondents who identified more as being Republican had lower perception levels, which underscores how preventive perceptions are a function of political party identification. The related findings on perceived severity and susceptibility are generally in line with prior research that documented that, as compared to liberals, conservatives have lower levels of fear of COVID-19 [39], with fear considered to be an outcome of elevated perceptions of severity and susceptibility [11]. However, arguably, what matters more is that preventive behaviors did not vary significantly by political party identification. This result differs from one prior county-level study that demonstrated that people in Republican counties were much less likely to practice social distancing than people in Democrat counties [20]. That preventive perceptions vary significantly by political party identification is indicative of the contemporary political divide in the United States and the polarization of news coverage in the partisan media [20,36,38]. While altering the predominant slant of news coverage in Fox News and related conservative media may be impossible, public health interventions that disseminate preventive information via other sources and help build media literacy to depoliticize health topics such as COVID-19 could help narrow the perceptual gap between Republicans and Democrats. Interestingly, though preventive perceptions vary by political party identification, preventive behaviors do not. Nevertheless, given the vast literature that indicates that perceptions drive behavior change, we believe that media can still play an important role in educating the public on politicized health topics such as COVID-19.

The results were less consistent when it came to the indicator of COVID-19 testing status, which was a measure of personal relevance [21]. Respondents who tested positive for COVID-19 had significantly higher perceptions of susceptibility in Study 1 and lower perceptions of severity in Study 2. The first result, which occurred very early in the pandemic when testing was not widely available, makes sense in that respondents who had tested positive would think they were more susceptible to COVID-19. In terms of the significant result in Study 2, it could be that respondents who tested positive for COVID-19 considered this infectious disease to be less severe because they did not experience major negative effects.

Limitations

Five limitations deserve acknowledgement. First, given this study's reliance on two cross-sectional data sets, no inferences of causation can be made. Second, self-report survey data have some measurement limitations, including social desirability concerns. Third, while the sample size of Study 2 (N=1037) is sufficient, the smaller sample size of Study 1 (N=274) may pose concerns in terms of elevated sampling error and attenuated statistical power. There is evidence of this in the regression results, where standardized coefficients of a certain size were deemed significant at *P* values of less than .05 in Table 4, but not in Table 2. For this reason, we referred to results that were nearing significance (ie, $P<.10$) in the reporting of results for Study 1. Fourth, dichotomous (ie, yes or no) questions, which have been used in prior research [13,14,41], provide a general picture of behavioral compliance, but not one of behavioral frequency. By using frequency measurement with a continuous scale for each preventive behavior, future research could depict how often millennials perform recommended COVID-19 preventive behaviors. Fifth, given changes across the pandemic in terms of COVID-19 risks and behavioral recommendations, we caution researchers in generalizing these results to other time frames across the pandemic, as well as to populations other than millennials. That said, this manuscript's use of two data sets helps mitigate some such concerns related to time frame.

Conclusions

In conclusion, this manuscript depicts US millennials' levels of COVID-19 preventive perceptions and behaviors, as well as their predictive factors. Understanding these levels and predictive factors has implications for theory, policy, and practical intervention. The analyses in both empirical studies highlight the importance of health information seeking in the face of a global pandemic, as well as the politicization of the COVID-19 pandemic. Future research should continue to investigate related processes in terms of COVID-19 and other pandemics. It could advance this manuscript's results by using panel survey data to derive inferences of causation and making assessments across later stages of the COVID-19 pandemic.

Conflicts of Interest

None declared.

References

1. Coronavirus Resource Center. COVID-19 Dashboard. Baltimore, MD: Johns Hopkins University & Medicine; 2021. URL: https://coronavirus.jhu.edu/map.html [accessed 2021-04-01]
2. Coronavirus pandemic (COVID-19). Our World in Data. 2020. URL: https://ourworldindata.org [accessed 2020-10-01]
3. The Department of Global Communications. UN tackles ‘infodemic’ of misinformation and cybercrime in COVID-19 crisis. United Nations. 2020 Mar 31. URL: https://www.un.org/en/un-coronavirus-communications-team/un-tackling-%E2%80%98infodemic%E2%80%99-misinformation-and-cybercrime-covid-19 [accessed 2020-09-01]
4. Wagner M, Macaya M, Yeung J, Renton A, Rahim Z, Isaac L. CDC director urges millennials to follow Covid-19 guidelines. CNN. 2020 Jun 26. URL: https://www.cnn.com/world/live-news/coronavirus-pandemic-06-26-20-intl/h_1ffa8c7ddf6e3b72fe232933ab2f3cb9 [accessed 2020-06-01]
5. Stone W. Younger adults are increasingly testing positive for the coronavirus. NPR. 2020 Jun 19. URL: https://www.npr.org/sections/health-shots/2020/06/19/880912184/younger-adults-are-increasingly-testing-positive-for-coronavirus [accessed 2020-11-01]
6. Lahut J. Trump coronavirus task force coordinator says millennials 'are the core group that will stop this virus'. Business Insider. 2020 Mar 16. URL: https://www.businessinsider.com/trump-dr-birx-calls-millennials-the-core-group-stop-coronavirus-2020-3 [accessed 2020-09-01]
7. Fry R. Millennials overtake Baby Boomers as America's largest generation. Pew Research Center. 2020 Apr 28. URL: https://www.pewresearch.org/fact-tank/2020/04/28/millennials-overtake-baby-boomers-as-americas-largest-generation/ [accessed 2020-09-01]
8. Fry R. Millennials are the largest generation in the US labor force. Pew Research Center. 2018 Apr 11. URL: https://www.pewresearch.org/fact-tank/2018/04/11/millennials-largest-generation-us-labor-force/ [accessed 2020-09-01]
9. Prensky M. Digital natives, digital immigrants Part 1. On the Horizon 2001 Sep;9(5):1-6. [doi: 10.1108/10748120110424816]
10. Kohut A, Taylor P, Keeter S, Parker K, Morin R, Cohn D, et al. Millennials: Confident. Connected. Open to Change. Washington, DC: Pew Research Center; 2010 Feb. URL: https://www.pewresearch.org/wp-content/uploads/sites/3/2010/10/millennials-confident-connected-open-to-change.pdf [accessed 2020-09-01]
11. Witte K. Putting the fear back into fear appeals: The extended parallel process model. Commun Monogr 1992;59(4):329-349. [doi: 10.1080/03637759209376276]
12. Skinner C, Tiro J, Champion V. The Health Belief Model. In: Glanz K, Rimer BK, Viswanath K, editors. Health Behaivor: Theory, Research, and Practice. 5th edition. San Francisco, CA: Jossey-Bass; 2015:75-94.
13. Camacho-Rivera M, Islam JY, Vidot DC. Associations between chronic health conditions and COVID-19 preventive behaviors among a nationally representative sample of US adults: An analysis of the COVID Impact Survey. Health Equity 2020;4(1):336-344 [FREE Full text] [doi: 10.1089/heq.2020.0031] [Medline: 32783017]
14. Breakwell GM, Fino E, Jaspal R. The COVID-19 Preventive Behaviors Index: Development and validation in two samples from the United Kingdom. Eval Health Prof 2021 Mar;44(1):77-86 [FREE Full text] [doi: 10.1177/0163278720983416] [Medline: 33430623]
15. Roberts JA, David ME. Improving predictions of COVID-19 preventive behavior: Development of a sequential mediation model. J Med Internet Res 2021 Mar 17;23(3):e23218 [FREE Full text] [doi: 10.2196/23218] [Medline: 33651707]
16. Rad RE, Mohseni S, Takhti HK, Azad MH, Shahabi N, Aghamolaei T, et al. Application of the protection motivation theory for predicting COVID-19 preventive behaviors in Hormozgan, Iran: A cross-sectional study. BMC Public Health 2021 Mar 08;21:1-11 [FREE Full text] [doi: 10.1186/s12889-021-10500-w] [Medline: 33685426]
17. Chan DKC, Zhang C, Weman-Josefsson K. Why people failed to adhere to COVID-19 preventive behaviors? Perspectives from an integrated behavior change model. Infect Control Hosp Epidemiol 2021 Mar;42(3):375-376 [FREE Full text] [doi: 10.1017/ice.2020.245] [Medline: 32408917]
18. Masters NB, Shih SF, Bukoff A, Akel KB, Kobayashi LC, Miller AL, et al. Social distancing in response to the novel coronavirus (COVID-19) in the United States. PLoS One 2020;15(9):e0239025 [FREE Full text] [doi: 10.1371/journal.pone.0239025] [Medline: 32915884]
19. Liu PL. COVID-19 information seeking on digital media and preventive behaviors: The mediation role of worry. Cyberpsychol Behav Soc Netw 2020 Oct;23(10):677-682. [doi: 10.1089/cyber.2020.0250] [Medline: 32498549]
20. Gollwitzer A, Martel C, Brady WJ, Pärnamets P, Freedman IG, Knowles ED, et al. Partisan differences in physical distancing are linked to health outcomes during the COVID-19 pandemic. Nat Hum Behav 2020 Nov;4(11):1186-1197. [doi: 10.1038/s41562-020-00977-7] [Medline: 33139897]
21. Johnson JD, Meischke H. A comprehensive model of cancer-related information seeking applied to magazines. Hum Commun Res 1993 Mar;19(3):343-367. [doi: 10.1111/j.1468-2958.1993.tb00305.x]
22. Lambert SD, Loiselle CG. Health information seeking behavior. Qual Health Res 2007 Oct;17(8):1006-1019. [doi: 10.1177/1049732307305199] [Medline: 17928475]

23. Yoon J, Huang H, Kim S. Trends in health information-seeking behaviour in the US foreign-born population based on the Health Information National Trends Survey, 2005 - 2014. Inf Res 2017;22(3):1-26 [FREE Full text] [doi: 10.32388/6p9glv]
24. National Cancer Institute. Health Information National Trends Survey (HINTS) 5, Cycle 1. Bethesda, MD: National Cancer Institute, Division of Cancer Control and Population Sciences; 2017. URL: https://hints.cancer.gov/dataset/HINTS-5_Cycle1_STATA.zip [accessed 2021-05-01]
25. National Cancer Institute. Health Information National Trends Survey (HINTS) 5, Cycle 3. Bethesda, MD: National Cancer Institute, Division of Cancer Control and Population Sciences; 2019. URL: https://hints.cancer.gov/dataset/HINTS5_Cycle3_Stata_20210305.zip [accessed 2021-05-01]
26. Zhao X, Fan J, Basnyat I, Hu B. Online health information seeking using "#COVID-19 Patient Seeking Help" on Weibo in Wuhan, China: Descriptive study. J Med Internet Res 2020 Oct 15;22(10):e22910 [FREE Full text] [doi: 10.2196/22910] [Medline: 33001838]
27. Jurkowitz M, Mitchell A. Older Americans continue to follow COVID-19 news more closely than younger adults. Pew Research Center. 2020 Apr 22. URL: https://www.journalism.org/2020/04/22/older-americans-continue-to-follow-covid-19-news-more-closely-than-younger-adults/ [accessed 2020-09-01]
28. Vogels EA, Perrin A, Rainie L, Anderson M. 53% of Americans say the internet has been essential during the COVID-19 outbreak. Pew Research Center. 2020 Apr 30. URL: https://www.pewresearch.org/internet/2020/04/30/53-of-americans-say-the-internet-has-been-essential-during-the-covid-19-outbreak/ [accessed 2020-09-01]
29. Beaudoin CE, Hong T. Health information seeking, diet and physical activity: An empirical assessment by medium and critical demographics. Int J Med Inform 2011;80(8):586-595. [doi: 10.1016/j.ijmedinf.2011.04.003] [Medline: 21640643]
30. Shim M, Kelly B, Hornik R. Cancer information scanning and seeking behavior is associated with knowledge, lifestyle choices, and screening. J Health Commun 2006 Feb;11 Suppl 1:157-172. [doi: 10.1080/10810730600637475] [Medline: 16641081]
31. Beaudoin CE, Tao CC. Modeling the impact of online cancer resources on the supporters of cancer patients. New Media Soc 2008;10(2):321-344. [doi: 10.1177/1461444807086477]
32. 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]
33. Paige SR, Miller MD, Krieger JL, Stellefson M, Cheong J. Electronic health literacy across the lifespan: Measurement invariance study. J Med Internet Res 2018 Jul 09;20(7):e10434 [FREE Full text] [doi: 10.2196/10434] [Medline: 29986848]
34. Oh YS. Predictors of self and surrogate online health information seeking in family caregivers to cancer survivors. Soc Work Health Care 2015;54(10):939-953. [doi: 10.1080/00981389.2015.1070780] [Medline: 26671245]
35. Party identification. Inter-university Consortium for Political and Social Research (ICPSR). Ann Arbor, MI: University of Michigan, Institute for Social Research; 2021. URL: https://www.icpsr.umich.edu/web/pages/instructors/setups/notes/party-identification.html [accessed 2021-05-01]
36. Political polarization in the American public. Pew Research Center. 2014 Jun 12. URL: https://www.pewresearch.org/politics/2014/06/12/political-polarization-in-the-american-public/ [accessed 2021-03-01]
37. Chen E, Chang H, Rao A, Lerman K, Cowan G, Ferrara E. COVID-19 misinformation and the 2020 US presidential election. HKS Misinf Rev 2021 Mar 3:1-17. [doi: 10.37016/mr-2020-57]
38. Hart PS, Chinn S, Soroka S. Politicization and polarization in COVID-19 news coverage. Sci Commun 2020 Aug 25;42(5):679-697 [FREE Full text] [doi: 10.1177/1075547020950735]
39. Winter T, Riordan BC, Pakpour AH, Griffiths MD, Mason A, Poulgrain JW, et al. Evaluation of the English version of the Fear of COVID-19 Scale and its relationship with behavior change and political beliefs. Int J Ment Health Addict 2020 Jun 15:1-11 [FREE Full text] [doi: 10.1007/s11469-020-00342-9] [Medline: 32837431]
40. Witte K, Meyer G, Martell D. Effective Health Risk Messages: A Step-By-Step Guide. Thousand Oaks, CA: SAGE Publications; 2001.
41. Kaiser Family Foundation. KFF Coronavirus Poll – March 2020. San Francisco, CA: Henry J Kaiser Family Foundation; 2020. URL: http://files.kff.org/attachment/Topline-KFF-Coronavirus-Poll.pdf [accessed 2020-09-01]
42. National Cancer Institute. Health Information National Trends Survey (HINTS) 4, Cycle 1. Bethesda, MD: National Cancer Institute, Division of Cancer Control and Population Sciences; 2012. URL: https://hints.cancer.gov/dataset/HINTS4_Cycle1_Stata.zip [accessed 2021-05-01]
43. Hampton KN. Social media and change in psychological distress over time: The role of social causation. J Comput Mediat Commun 2019;24(5):205-222. [doi: 10.1093/jcmc/zmz010]
44. Allison P. When can you safely ignore multicollinearity? Statistical Horizons. 2012 Sep 10. URL: https://statisticalhorizons.com/multicollinearity [accessed 2021-02-01]
45. Lee JJ, Kang K, Wang MP, Zhao SZ, Wong JYH, O'Connor S, et al. Associations between COVID-19 misinformation exposure and belief with COVID-19 knowledge and preventive behaviors: Cross-sectional online study. J Med Internet Res 2020 Nov 13;22(11):e22205 [FREE Full text] [doi: 10.2196/22205] [Medline: 33048825]
46. Twenge JM, Campbell SM. Generational differences in psychological traits and their impact on the workplace. J Manage Psychol 2008 Nov 07;23(8):862-877. [doi: 10.1108/02683940810904367]

Abbreviations

EPPM:	Extended Parallel Process Model
H1:	Hypothesis 1
H2:	Hypothesis 2
HINTS:	Health Information National Trends Survey
HISB:	health information–seeking behavior
OLS:	ordinary least squares
RQ1:	Research Question 1
RQ2:	Research Question 2
VIF:	variance inflation factor

Edited by C Basch; submitted 21.05.21; peer-reviewed by L Suppan, MA Bahrami; comments to author 31.05.21; revised version received 15.06.21; accepted 19.06.21; published 12.08.21.

Please cite as:
Beaudoin CE, Hong T
Predictors of COVID-19 Preventive Perceptions and Behaviors Among Millennials: Two Cross-sectional Survey Studies
J Med Internet Res 2021;23(8):e30612
URL: https://www.jmir.org/2021/8/e30612
doi:10.2196/30612
PMID:34182460

©Christopher E Beaudoin, Traci Hong. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 12.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Digital Surveillance Through an Online Decision Support Tool for COVID-19 Over One Year of the Pandemic in Italy: Observational Study

Alberto Eugenio Tozzi[1], MD; Francesco Gesualdo[1], MD; Emanuele Urbani[2], MD; Alessandro Sbenaglia[2], MSc; Roberto Ascione[3], MSc; Nicola Procopio[3], MSc; Ileana Croci[1], DStat; Caterina Rizzo[4], MD

[1]Multifactorial and Complex Diseases Research Area, Bambino Gesù Children's Hospital IRCCS, Rome, Italy

[2]Paginemediche, Salerno, Italy

[3]Healthware Group, Salerno, Italy

[4]Clinical Pathways and Epidemiology Unit, Bambino Gesù Children's Hospital IRCCS, Rome, Italy

Corresponding Author:
Francesco Gesualdo, MD
Multifactorial and Complex Diseases Research Area
Bambino Gesù Children's Hospital IRCCS
Piazza S. Onofrio, 4
Rome, 00165
Italy
Phone: 39 3388437474
Email: francesco.gesualdo@opbg.net

Abstract

Background: Italy has experienced severe consequences (ie, hospitalizations and deaths) during the COVID-19 pandemic. Online decision support systems (DSS) and self-triage applications have been used in several settings to supplement health authority recommendations to prevent and manage COVID-19. A digital Italian health tech startup, Paginemediche, developed a noncommercial, online DSS with a chat user interface to assist individuals in Italy manage their potential exposure to COVID-19 and interpret their symptoms since early in the pandemic.

Objective: This study aimed to compare the trend in online DSS sessions with that of COVID-19 cases reported by the national health surveillance system in Italy, from February 2020 to March 2021.

Methods: We compared the number of sessions by users with a COVID-19–positive contact and users with COVID-19–compatible symptoms with the number of cases reported by the national surveillance system. To calculate the distance between the time series, we used the dynamic time warping algorithm. We applied Symbolic Aggregate approXimation (SAX) encoding to the time series in 1-week periods. We calculated the Hamming distance between the SAX strings. We shifted time series of online DSS sessions 1 week ahead. We measured the improvement in Hamming distance to verify the hypothesis that online DSS sessions anticipate the trends in cases reported to the official surveillance system.

Results: We analyzed 75,557 sessions in the online DSS; 65,207 were sessions by symptomatic users, while 19,062 were by contacts of individuals with COVID-19. The highest number of online DSS sessions was recorded early in the pandemic. Second and third peaks were observed in October 2020 and March 2021, respectively, preceding the surge in notified COVID-19 cases by approximately 1 week. The distance between sessions by users with COVID-19 contacts and reported cases calculated by dynamic time warping was 61.23; the distance between sessions by symptomatic users was 93.72. The time series of users with a COVID-19 contact was more consistent with the trend in confirmed cases. With the 1-week shift, the Hamming distance between the time series of sessions by users with a COVID-19 contact and reported cases improved from 0.49 to 0.46. We repeated the analysis, restricting the time window to between July 2020 and December 2020. The corresponding Hamming distance was 0.16 before and improved to 0.08 after the time shift.

Conclusions: Temporal trends in the number of online COVID-19 DSS sessions may precede the trend in reported COVID-19 cases through traditional surveillance. The trends in sessions by users with a contact with COVID-19 may better predict reported cases of COVID-19 than sessions by symptomatic users. Data from online DSS may represent a useful supplement to traditional surveillance and support the identification of early warning signals in the COVID-19 pandemic.

(J Med Internet Res 2021;23(8):e29556) doi:10.2196/29556

KEYWORDS

COVID-19; public health; surveillance; digital surveillance; internet; online decision support system; decision support; support; online tool; Italy; observational

Introduction

As of March 2021, the World Health Organization had estimated that almost 120 million cases of COVID-19 had occurred worldwide, with more than 2.5 million deaths [1]. The first cases of COVID-19 in Europe date back to January 2020, with the first 2 cases identified in Italy on January 31 in 2 Chinese tourists [2]. Since then, Italy has been severely hit by the pandemic [3]. As in other countries, from the earliest phases of the pandemic, the Italian Ministry of Health developed an information campaign with recommendations for managing contacts of SARS-CoV-2–positive individuals and information on symptoms that could suggest a possible diagnosis of COVID-19.

Despite the availability of these recommendations in the media, the surge of requests for information and advice could not be entirely managed by the health authorities through public health care helplines and family doctors. Indeed, a large part of the population tried to find appropriate answers to their questions on the web [4].

Online decision support tools have been used in the past, such as during the A/H1N1pdm09 pandemic [5], to meet the information needs of the public at large, improve measures to prevent transmission, and support the management of symptomatic cases. During the COVID-19 pandemic, online decision support systems (DSS) and artificial intelligence chatbots have also been used for digital triage and self-diagnosis and have been shown to be accurate in identifying suspected COVID-19 cases and efficient in decompressing the pressure placed on emergency rooms and helplines [6-12]. On the other hand, the adoption of digital tools encounters barriers and limitations intrinsic to their novelty, the digital divide, and other local circumstances [6,9].

To provide clinical and public health recommendations in line with the evolving guidelines issued by the Italian Ministry of Health, in February 2020, Paginemediche, a digital Italian health tech startup, developed a noncommercial, online DSS available in Italian [13]. The system was designed as a simple algorithm for assisting individuals manage their potential exposure to COVID-19 and has a chat user interface. The system was available on several web pages including landing pages of the regional health systems and was advertised through social media. A link to the online DSS was also hosted on the home page of the Regional Health Authorities of Lombardy and Campania and of Trento autonomous province.

Digital platforms collecting self-reported information have been shown to provide complementary information to traditional epidemiological surveillance [14]. In particular, it has been reported that patterns of use of online tools for COVID-19 self-triage may be considered a proxy of confirmed cases reported to national surveillance systems and may provide early signals of COVID-19 cases [15].

As our online DSS was available on a national scale from the early phase of the pandemic and for a time period of 1 year, we examined it as a potential data source for digital surveillance of COVID-19 in Italy. Our hypothesis was that trends in the use of the online DSS may precede those observed in the national surveillance system, as symptomatic individuals and those who have been in contact with a COVID-19 case may access the online DSS before undergoing a laboratory test for confirmation.

The aim of the present study was to describe the general characteristics of users of the online DSS; compare the trends in online DSS sessions with those of COVID-19 cases reported by the national health surveillance system in Italy, from February 2020 to March 2021; and study the time lag between the online DSS and the surveillance system data.

Methods

Our online DSS was developed as a web-based algorithm to triage users with COVID-19–compatible symptoms or with a history of contact with a confirmed COVID-19 case and to provide recommendations on appropriate management (including testing), according to guidelines from the Italian Ministry of Health [16]. Anonymous access to the system is open to any user, and interaction is based on a chat interface. The system is available on the Paginemediche website [13].

Briefly, at the beginning of the session, the system collects information on age group and place of residence. Then, the user is asked to select the reason for accessing the system from among the following: (1) evaluation of symptoms, (2) contact with a confirmed COVID-19 case, (3) notification received from the national contact tracing app [17], (4) positive swab for COVID-19. The system then proceeds by matching the information provided by the users with national guidelines and provides tailored recommendations, such as asking for immediate medical support, contacting the family doctor, performing a diagnostic test, or applying nonpharmaceutical measures such as isolation and self-quarantine.

As the purpose of this study was to test if access to the online DSS preceded (and therefore was able to predict) national surveillance system trends, we restricted our analysis to users accessing the system for the evaluation of symptoms (SYM) or for contact with a confirmed COVID-19 case (CON).

We analyzed the information on users' age and location and the symptoms reported by age group using descriptive statistics. We then explored the temporal trends in SYM and CON sessions, and we compared them with trends in COVID-19 cases officially reported by the Italian COVID-19 national surveillance system, which is based on the data from laboratory-confirmed

SARS-CoV-2 infections provided daily by regional health authorities to the National Health Institute. To perform the comparison, we first used moving averages to display the curves.

where p_i are data points.

K was set at a value of 7, as data from the national surveillance system were published weekly.

We then scaled each time series in a range between 0 and 1, applying the following formula:

To calculate the distance between the time series of cases reported through the national surveillance system and those of SYM and CON, we used the dynamic time warping algorithm (DTW) [18,19]. DTW measures similarity between 2 temporal sequences, which may vary in speed. The sequences are "warped" nonlinearly in the time dimension to determine a measure of their similarity, independent of nonlinear variations in time.

To confirm the hypothesis that the online DSS trends anticipated those of reported COVID-19 cases, we first applied Symbolic Aggregate approXimation (SAX) encoding [20] to the time series in 1-week periods. Then, we calculated the Hamming distance [21,22] between the SAX strings. SAX transforms a time series into a sequence of symbols that represents a range of values, allowing the application of the dimensionality reduction on the time series [20,23]. SAX encoding was chosen as it allows dimensionality/numerosity reduction of the time series and as it is relatively easy to understand and compute. The dimensionality reduction was based on Piecewise Aggregate Approximation [20,22]. To verify that the online DSS anticipated the trends observed in notified cases, we shifted its time series 1 week ahead, and we measured the improvement in the Hamming distance [24]. The 1-week shift was chosen empirically based on the knowledge that the average incubation period is 4 days [25], the median time to diagnosis is 5.85 days [26], and most Italian laboratories have been providing results for PCR tests 24-48 hours after the test. Other authors have chosen the same time shift for modelling purposes [27].

This study includes data collected by the online DSS from February 25, 2020 to March 15, 2021 and the corresponding figures of COVID-19 cases in Italy published by the national COVID-19 surveillance system in their open access repository [28].

The statistical analysis was conducted using Python 3.8.5.

Considering the nature of the analysis, the current study did not require approval by the local ethics committee according to current legislation. However, a notification including the study characteristics was sent to the Bambino Gesù Ethical Committee.

Results

We recorded a total of 75,557 sessions in the online DSS during the study period. Among them, 65,207 were SYM, and 19,062 were CON. Among the 65,207 sessions of users with symptoms, 8692 stated they also had contact with a COVID-19–positive individual.

Age was missing in nearly 32% (24,084/75,557, 31.87%) of the sessions, and place of residence was missing in 28.27% (21,360/75,557), probably because the input of information about age and place of residence was not compulsory in the first 2 months of activity in the online DSS. Among the 51,473 sessions for which age was recorded, the majority (40,923/51,473, 79.50%) were by adults in the 19–64-year-old age group: 19-24 years old, 4830/51,473, 9.38%; 25-34 years old, 9054/51,473, 17.59%; 35-44 years old, 10,383/51,473, 20.17%; 45-54 years old, 9852/51,473, 19.14%; 55-64 years old, 6804/51,473, 13.22%. Users 65 years old and older represented 12.89% (6636/51,473) of the sessions (65-74 years, 4339/51,473, 8.43%; ≥75 years, 2297/51,473, 4.46%), and 7.60% (3914/51,473) of the sessions were by users ≤18 years of age. Sessions were established from all Italian regions, at a proportion between 0.3 and 0.6 per 1000 inhabitants. The number of sessions by region are described in Figure 1. Three regions were more represented, reflecting more active local endorsement by regional authorities: 2 in Northern Italy, namely Lombardy and the Trento autonomous province, and 1 in Southern Italy, namely Campania.

Symptoms reported by online DSS users by age group are illustrated in Table 1.

Symptoms reported by users varied with age. In all age groups, fever was the most reported symptom, although the proportion of patients reporting fever was higher in younger individuals and decreased with age. A similar trend was observed for respiratory symptoms, such as cold and sore throat. Among other symptoms, myalgia, low blood pressure, and dizziness were most frequently reported among people belonging to older age groups.

Figure 1. Number of sessions by the number of residents in each region.

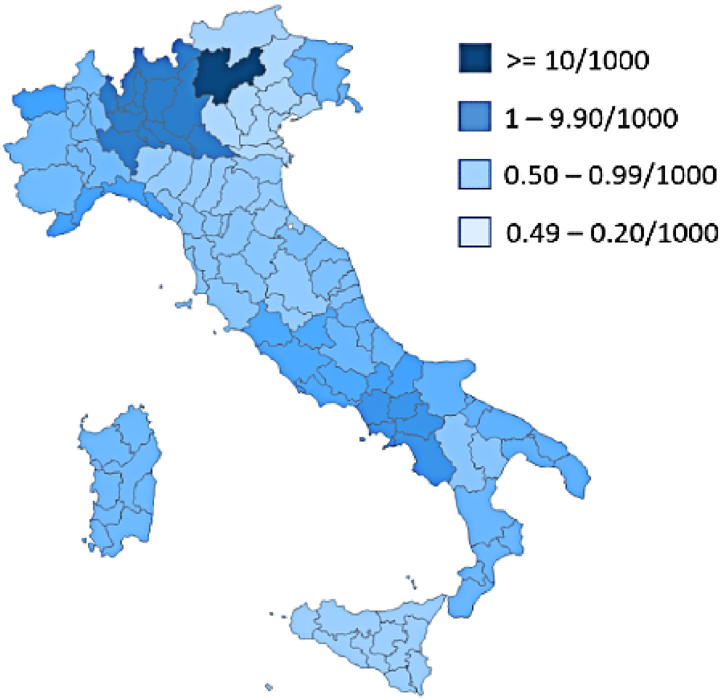


Table 1. Distribution of symptoms reported in each session by age group (n=51,473).

Symptom	Age (years), n (%)			
	0-18 (n=3914)	19-64 (n=40,923)	65-74 (n=4339)	≥75 (n=2297)
Fever	1912 (48.85)	12,896 (31.51)	993 (22.89)	625 (27.21)
Cold	1489 (38.04)	9584 (23.42)	801 (18.46)	406 (17.68)
Sore throat	1027 (26.24)	10,264 (25.08)	628 (14.47)	305 (13.28)
Cough	881 (22.51)	8540 (20.87)	831 (19.15)	441 (19.20)
Headache	437 (11.17)	4870 (11.90)	280 (6.45)	82 (3.57)
Myalgia	262 (6.69)	4947 (12.09)	643 (14.82)	284 (12.36)
Dispnea	381 (9.73)	4296 (10.50)	362 (8.34)	314 (13.67)
Low blood pressure	93 (2.38)	2612 (6.38)	611 (14.08)	322 (14.02)
Dizziness	301 (7.69)	2504 (6.12)	298 (6.87)	251 (10.93)
Drowsiness	214 (5.47)	2195 (5.36)	269 (6.20)	220 (9.58)
Tachycardia	260 (6.64)	2043 (4.99)	265 (6.11)	138 (6.01)
Nausea	285 (7.28)	1947 (4.76)	157 (3.62)	61 (2.66)

Figure 2 shows the numbers of SYM and CON sessions over time and the number of confirmed COVID-19 cases officially reported to the national surveillance system.

The highest number of sessions was recorded in the early phases of the pandemic, when COVID-19–confirmed cases started to increase very quickly.

A less pronounced peak was observed in the number of CON sessions in the same period. A second peak, both in SYM and CON sessions, was observed during a large second wave of COVID-19 cases, which, in Italy, started in October 2020. Finally, a third peak was observed in March 2021, when Italy experienced a third surge of cases. Figure 3 shows the trends in SYM sessions, CON sessions, and COVID-19 cases as moving averages in 1-week periods and scaling values between 0 and 1.

The 3 time series in the entire time period had a similar trend, independently from the amplitude of the peaks, although an increase in SYM and CON anticipated the trend in confirmed COVID-19 cases by approximately 1 week. The distance between SYM and notified COVID-19 cases calculated through DTW (the lower the distance, the higher the similarity between the curves) was 93.72, while that between CON and notified COVID-19 cases was 61.23.

As CON was more consistent with the trend in confirmed cases than SYM, we tested the hypothesis that this series anticipated confirmed cases by 1 week. After applying SAX encoding, we measured the Hamming distance between the time series of confirmed COVID-19 cases and the CON series (1) as recorded by the system and (2) shifted by 1 week. After applying the 1-week shift, the Hamming distance improved from 0.49 to 0.46. We also repeated the analysis restricting the time window to the time period between July 2020 and December 2020. The corresponding Hamming distance was 0.16 before shifting the time series and improved to 0.08 after the time shift (Figure 4).

Figure 2. Weekly number of COVID-19 confirmed cases (red), SYM (symptomatic users) sessions (yellow), and CON (users with contact with a COVID-19 case; green) sessions between February 2020 and March 2021.

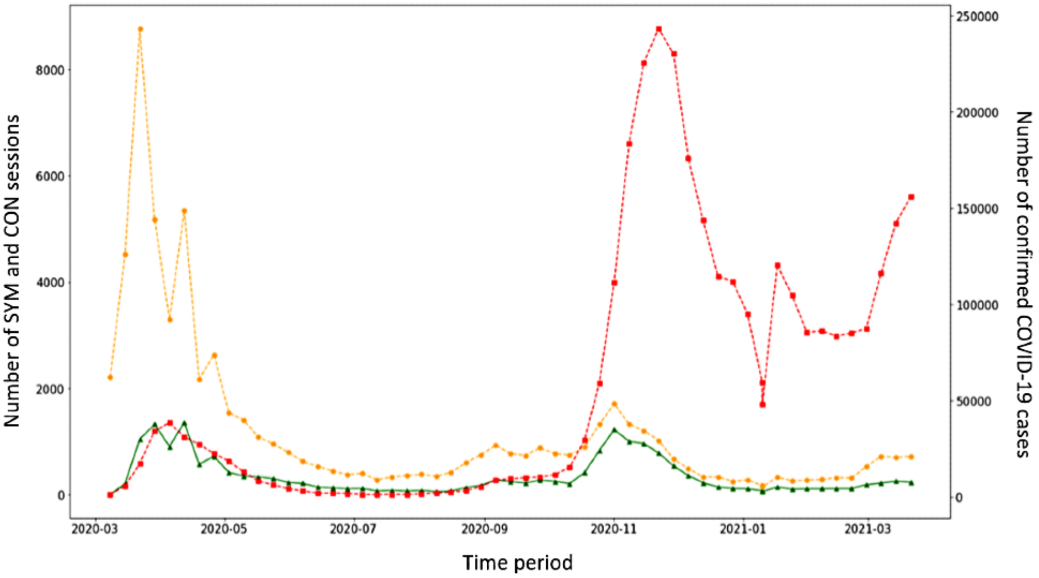


Figure 3. Time trends in SYM (symptomatic users) sessions (yellow), CON (users with contact with a COVID-19 case) sessions (green), and COVID-19 cases (red). Values are illustrated as rolling means in 1-week periods and scaled to values between 0 and 1.

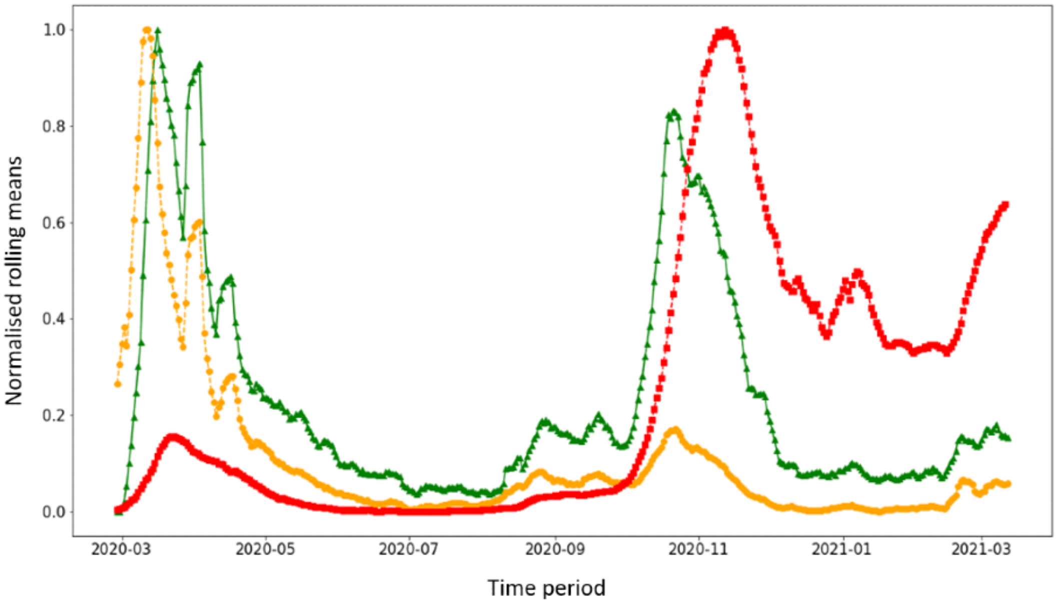
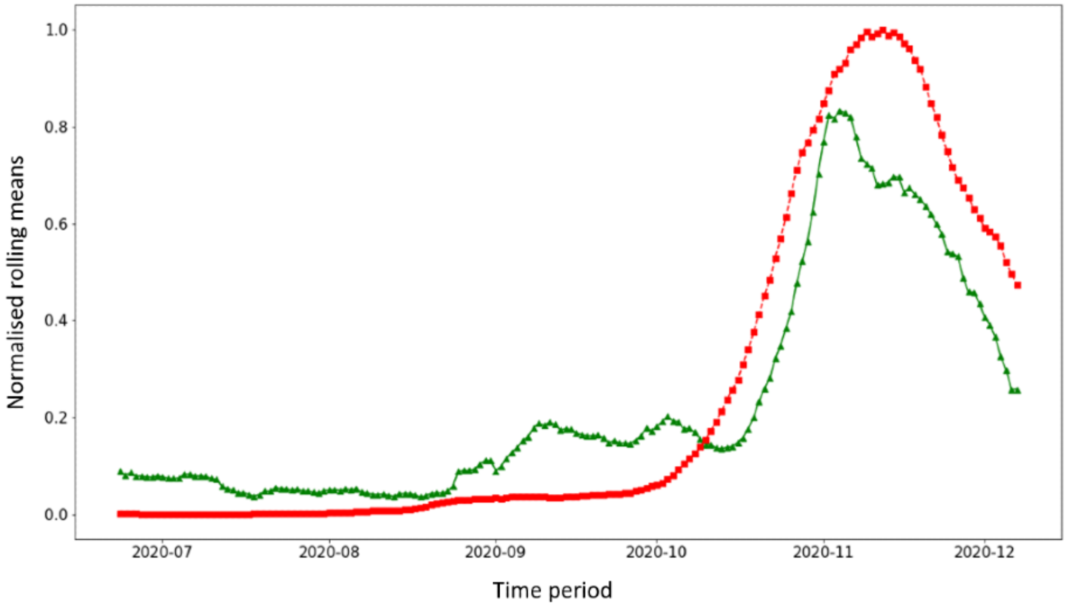


Figure 4. Time trends in CON (users with contact with a COVID-19 case) sessions and notified COVID-19 cases, Italy, July 2020 to December 2020. Values are illustrated as rolling means and scaled from 0 to 1. The CON time series is shifted 7 days ahead.



Discussion

In this study, we showed that users of a national online DSS reporting contact with a COVID-19 case anticipated trends observed by the Italian national surveillance system by 1 week.

Surveillance systems for infectious diseases frequently suffer from lack of timeliness and difficulties in case finding that may delay the implementation of effective, data-based preventative strategies [29,30]. Digital surveillance has been indicated as a potential supplement to traditional surveillance due to the fact that web and social media users may spontaneously leave online traces related to their health conditions, which can be detected by digital surveillance systems in a timely manner [31].

Classically, digital traces for surveillance are classified as either based on the public's demand of information (eg, volume of searches on Google) or on information supplied by internet and social media users (eg, tweets reporting symptoms of a specific disease) [32]. Indeed, several attempts have been made to track and interpret the signals from queries on search engines or from keywords or symptoms reported on social media to promptly recognize the emergence of infectious diseases, with mixed success [33]. As a matter of fact, trends in queries on search engines are strongly affected by media coverage that may not accurately reflect the epidemic trends [34].

With this study, we explored the potential of a digital surveillance system based on a mixed demand-supply method. Users accessed our online DSS with a specific request for information and were invited to provide specific data on their condition through a structured framework. The algorithm guided the user through predefined questions based on national guidelines for COVID-19 prevention and contact management, with the aim to provide tailored, actionable recommendations.

Our results show that temporal trends in the number of sessions on an online COVID-19 DSS dedicated to the general public may precede the trend in confirmed COVID-19 cases obtained through the national surveillance system by 7 days. We found a high correlation between the trend in sessions by users accessing the online DSS (with either a COVID-19 contact or symptoms) and the number of notified COVID-19 cases, documented by the short distance between the trends as estimated with DTW. However, the trends in sessions by users who had contact with COVID-19 cases predicted the trend in COVID-19 cases better than sessions by users with symptoms. This observation is plausible as, on the one hand, recognition of contact with someone with COVID-19 usually precedes the onset of symptoms, and, on the other hand, contact with a COVID-19 case can be considered as a more specific proxy of COVID-19, as the clinical picture of COVID-19 may overlap with that caused by other diseases. The correlation between sessions by users with a COVID-19 contact and confirmed cases greatly improved when the analysis was restricted to the central phase of the pandemic, from July 2020 to December 2020. In this time period, Italy experienced the highest surge of COVID-19 cases.

To our knowledge, only one other study attempted to use sessions in an online DSS as a proxy for case notification [15], showing lead times of 3 days in China and 19 days in the United States. Although the recognition of contact with a COVID-19 case may be delayed, our estimate of a 7-day lead time from users with contact with a COVID-19 case is consistent with the incubation of COVID-19 [26] and the time needed to perform a diagnostic test and obtain results.

These results show that the utility of an online DSS dedicated to the general public may go beyond its original purpose of guiding the public through recommendations for the management of cases or contacts.

First, these systems may generate early warning signals that may inform early interventions (eg, tailored restriction measures), as is desirable during the current pandemic. Data from our system may be particularly valuable in predicting large and rapid increases in COVID-19 cases, as shown by the higher correlation between the trend in contacts and national notifications.

Second, the high correlation estimates obtained through DTW support the inclusion of online DSS–based data in the development of prediction models, which could be useful to anticipate an increase in cases as well as to predict the timing of a reduction in cases. This would support national or regional institutions in more finely programming (timing, intensity, logistics, economic impact) the restrictive measures to contain the epidemic.

Digital data have already been used in prediction models for influenza. Through the analysis of the volume of online searches for flu-related terms, Google Flutrends offered a forecast of the trend in influenza-like illness cases in the United States [35], and a similar approach has recently been used in the Netherlands [36]. Data for participatory surveillance have also been integrated with classic surveillance data [37]. Although prediction models including participatory surveillance data can suffer from a lack of representativeness of the population sample, modeling and simulation can help control this bias [38].

In a recent study, Kogan et al [39] attempted to combine and harmonize different digital data streams to predict COVID-19 trends. They included Google Trends, Twitter, UpToDate, and mobility data, but no data from online DSSs were taken into account, which could have potentially improved the performance of the model.

For the purpose of informing prediction models with digital surveillance data, online DSS data would benefit from integration with data from contact tracing apps. Contact tracing apps have been developed in several countries, including Italy, to support public health measures for COVID-19 containment. These digital tools send a notification of exposure to a COVID-19–positive case to contacts and recommend preventative actions. However, data about the number of contacts and other personal characteristics remain stored in personal smartphones and cannot be used as a data source for surveillance. Moreover, although Italy has been considered as one of the countries with the highest acceptance of apps for contact tracing [40], only 20% of the general population actually downloaded the app as of March 2021 [41].

In this article, we report national data obtained through our online DSS. However, it should be noted that a very high user rate has been recorded in the Lombardia and Campania regions and in the autonomous province of Trento. The potential of predicting actual COVID-19 trends was more accurate in areas with a higher coverage of the system. In these regions, the online DSS had been actively promoted by local health institutions. The involvement of health authorities and the promotion of these systems by digital marketing through online social media are likely to increase the use of these digital tools, underlying their value in providing tailored recommendations to individuals. Stronger involvement of health authorities, by increasing participation and accuracy, would also allow us to develop prediction models on a regional — rather than on a national — basis.

This study has several strengths. First, data were available over a 12-month period on a national scale. We also found that the online DSS was accessed from all Italian regions, which was reassuring for the representativeness of the data sample. The system was also open to any user, as there was no age restriction for accessing the system.

The study also has obvious limitations. The most important is that we could not validate information provided by users, as expected from any anonymous online tool. Moreover, the number of sessions in an online DSS like ours may be affected by different factors that may increase the background noise and make the interpretation of trends difficult. Some of the fluctuations observed in the use of the online DSS may reflect a different perception of risk during different phases of the pandemic. Indeed, infodemics may affect perception of risk of infection that may result in an increase in online searches for information by the general public [42]. We cannot exclude that users' access to our online DSS might have been biased by different trends in interest by the general public, possibly affected by different media coverage of the pandemic throughout the study period. Nevertheless, the high correlation between our access data and the epidemic curve suggests that the variation of users' interest and, subsequently, of their online searches has not had a major impact on the use of our online DSS, which seems to have been primarily affected by the disease incidence. Nevertheless, in order to better validate a system like the one proposed in our study, future research should better estimate the impact of online searches on access to the online DSS, possibly correcting the analysis for the volume of searches based on keywords related to the epidemic.

Moreover, especially in the early phase of the pandemic, users may have tried to simulate different scenarios to better understand what to expect in case of symptoms or contact with a COVID-19 case. As a matter of fact, when restricting the analysis by excluding the earliest phase of the pandemic, the consistency of the online DSS time series with confirmed COVID-19 cases was higher. Furthermore, we did not collect socioeconomic information, and we could not attempt to adjust our analyses for confounders. Finally, all data streams from digital systems may suffer from selection bias due to the digital divide [43]. As a matter of fact, only 13% of the sessions were by individuals over the age of 65 years, which was the age group mostly affected in the first phase of the pandemic. Therefore, we should consider that this age group may have not been well represented in the online DSS sessions, and this might have affected the accuracy of our prediction. Nevertheless, we were not interested in estimating the incidence by age group based on the online DSS data, as our main objective was to study the trends in access to the online DSS and their relationship with official surveillance data. The high number of sessions recorded by any age group and from any region and their consistency over time with the trend in notified cases are reassuring that potential major biases are not present.

In conclusion, as online DSSs like the one described in this study are widely available and may be completely anonymous, they may represent a significant source of data to inform public health actions, may supplement other epidemiological information collected through other means, and may help to develop more accurate prediction models for pandemic trends.

Public health strategies for immunization and containment are usually based on incidence of preventable diseases and vaccine coverage, which carry some latency due to the notification systems. Despite the observed limitations, prediction models based on systems like the one described in our study may anticipate other epidemiological signals of epidemic surge and, when integrated with official surveillance data, can support timely decisions for mitigating the pandemic spread.

Digital tools for managing COVID-19 not only have epidemiological implications but also may be extremely helpful in disseminating appropriate recommendations to the general population in a universal strategy.

Although it remains to be assessed how frequently recommendations provided are actually followed by users and selection bias should be better understood and addressed, these tools represent a potential data source for public health that may help, in combination with others, to expeditiously implement containment strategies.

Acknowledgments

We thank Kiersten Miller for revising the English manuscript and Tina Dell'Armi for providing support for data management and analysis.

Authors' Contributions

AET conceived the study and drafted the manuscript. RA conceived the study and contributed to interpretation of results. FG drafted the manuscript and contributed to interpretation of results. CR contributed to interpretation of results and revised the manuscript. EU and AS contributed to data extraction, management, and analysis. NP and IC performed the statistical analysis.

Conflicts of Interest

EU, AS, RA, and NP work for Paginemediche and Healthware Group, which were involved in the development of the online DSS included on the Paginemediche website. CR received grants for participating in Advisory Boards. The other authors declare no conflicts of interest.

References

1. WHO Coronavirus (COVID-19) Dashboard. World Health Organization. URL: https://covid19.who.int [accessed 2021-08-01]
2. Spiteri G, Fielding J, Diercke M, Campese C, Enouf V, Gaymard A, et al. First cases of coronavirus disease 2019 (COVID-19) in the WHO European Region, 24 January to 21 February 2020. Euro Surveill 2020 Mar;25(9):1 [FREE Full text] [doi: 10.2807/1560-7917.ES.2020.25.9.2000178] [Medline: 32156327]
3. Onder G, Rezza G, Brusaferro S. Case-Fatality Rate and Characteristics of Patients Dying in Relation to COVID-19 in Italy. JAMA 2020 May 12;323(18):1775-1776. [doi: 10.1001/jama.2020.4683] [Medline: 32203977]
4. Rovetta A, Bhagavathula AS. COVID-19-Related Web Search Behaviors and Infodemic Attitudes in Italy: Infodemiological Study. JMIR Public Health Surveill 2020 May 05;6(2):e19374 [FREE Full text] [doi: 10.2196/19374] [Medline: 32338613]
5. Spaulding AB, Radi D, Macleod H, Lynfield R, Larson M, Hyduke T, et al. Design and implementation of a statewide influenza nurse triage line in response to pandemic H1N1 influenza. Public Health Rep 2012;127(5):532-540 [FREE Full text] [doi: 10.1177/003335491212700509] [Medline: 22942472]
6. Lai L, Wittbold KA, Dadabhoy FZ, Sato R, Landman AB, Schwamm LH, et al. Digital triage: Novel strategies for population health management in response to the COVID-19 pandemic. Healthc (Amst) 2020 Dec;8(4):100493 [FREE Full text] [doi: 10.1016/j.hjdsi.2020.100493] [Medline: 33129176]
7. Hautz WE, Exadaktylos A, Sauter TC. Online forward triage during the COVID-19 outbreak. Emerg Med J 2020 Dec 11:1 [FREE Full text] [doi: 10.1136/emermed-2020-209792] [Medline: 33310732]
8. Haase CB, Bearman M, Brodersen J, Hoeyer K, Risor T. 'You should see a doctor', said the robot: Reflections on a digital diagnostic device in a pandemic age. Scand J Public Health 2021 Feb;49(1):33-36 [FREE Full text] [doi: 10.1177/1403494820980268] [Medline: 33339468]
9. Fan X, Chao D, Zhang Z, Wang D, Li X, Tian F. Utilization of Self-Diagnosis Health Chatbots in Real-World Settings: Case Study. J Med Internet Res 2021 Jan 06;23(1):e19928 [FREE Full text] [doi: 10.2196/19928] [Medline: 33404508]
10. Judson TJ, Odisho AY, Young JJ, Bigazzi O, Steuer D, Gonzales R, et al. Implementation of a digital chatbot to screen health system employees during the COVID-19 pandemic. J Am Med Inform Assoc 2020 Jul 01;27(9):1450-1455 [FREE Full text] [doi: 10.1093/jamia/ocaa130] [Medline: 32531066]
11. Galmiche S, Rahbe E, Fontanet A, Dinh A, Bénézit F, Lescure F, et al. Implementation of a Self-Triage Web Application for Suspected COVID-19 and Its Impact on Emergency Call Centers: Observational Study. J Med Internet Res 2020 Nov 23;22(11):e22924 [FREE Full text] [doi: 10.2196/22924] [Medline: 33147165]
12. Martin A, Nateqi J, Gruarin S, Munsch N, Abdarahmane I, Zobel M, et al. An artificial intelligence-based first-line defence against COVID-19: digitally screening citizens for risks via a chatbot. Sci Rep 2020 Nov 04;10(1):19012 [FREE Full text] [doi: 10.1038/s41598-020-75912-x] [Medline: 33149198]

13. Paginemediche: la piattaforma che connette medici e pazienti. URL: https://www.paginemediche.it/ [accessed 2021-08-01]
14. Menni C, Valdes AM, Freidin MB, Sudre CH, Nguyen LH, Drew DA, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. Nat Med 2020 Jul;26(7):1037-1040 [FREE Full text] [doi: 10.1038/s41591-020-0916-2] [Medline: 32393804]
15. Hswen Y, Brownstein JS, Xu X, Yom-Tov E. Early detection of COVID-19 in China and the USA: summary of the implementation of a digital decision-support and disease surveillance tool. BMJ Open 2020 Dec 10;10(12):e041004 [FREE Full text] [doi: 10.1136/bmjopen-2020-041004] [Medline: 33303453]
16. Cosa sapere su test, tracciamento, quarantena. Ministero della Salute. URL: https://www.salute.gov.it/portale/nuovocoronavirus/dettaglioFaqNuovoCoronavirus.jsp?lingua=italiano&id=244 [accessed 2021-08-01]
17. immuni-app / immuni-documentation. GitHub. URL: https://github.com/immuni-app/immuni-documentation [accessed 2021-08-01]
18. Sakoe H, Chiba S. Dynamic programming algorithm optimization for spoken word recognition. IEEE Trans. Acoust., Speech, Signal Process 1978 Feb;26(1):43-49. [doi: 10.1109/TASSP.1978.1163055]
19. Salvador S, Chan P. Toward accurate dynamic time warping in linear time and space. IDA 2007 Oct 10;11(5):561-580. [doi: 10.3233/ida-2007-11508]
20. Lin J, Keogh E, Wei L, Lonardi S. Experiencing SAX: a novel symbolic representation of time series. Data Min Knowl Disc 2007 Apr 3;15(2):107-144. [doi: 10.1007/s10618-007-0064-z]
21. Hamming RW. Error Detecting and Error Correcting Codes. Bell System Technical Journal 1950;29(2):147-160 [FREE Full text] [doi: 10.1002/j.1538-7305.1950.tb00463.x]
22. Keogh E, Chakrabarti K, Pazzani M, Mehrotra S. Dimensionality Reduction for Fast Similarity Search in Large Time Series Databases. Knowledge and Information Systems 2001 Aug;3(3):263-286. [doi: 10.1007/pl00011669]
23. Lin J, Keogh E, Lonardi S, Chiu B. A symbolic representation of time series, with implications for streaming algorithms. 2003 Presented at: 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery; June 13, 2003; San Diego, CA. [doi: 10.1145/882082.882086]
24. scipy.spatial.distance.hamming. SciPy. URL: https://docs.scipy.org/doc/scipy/reference/generated/scipy.spatial.distance.hamming.html [accessed 2021-08-01]
25. Zardini A, Galli M, Tirani M, Cereda D, Manica M, Trentini F, et al. A quantitative assessment of epidemiological parameters to model COVID- 19 burden. Cornell University. 2021 Mar 29. URL: http://arxiv.org/abs/2103.15780 [accessed 2021-08-01]
26. Ejima K, Kim KS, Ludema C, Bento AI, Iwanami S, Fujita Y, et al. Estimation of the incubation period of COVID-19 using viral load data. Epidemics 2021 Jun;35:100454 [FREE Full text] [doi: 10.1016/j.epidem.2021.100454] [Medline: 33773195]
27. Turk PJ, Tran TP, Rose GA, McWilliams A. A predictive internet-based model for COVID-19 hospitalization census. Sci Rep 2021 Mar 03;11(1):5106 [FREE Full text] [doi: 10.1038/s41598-021-84091-2] [Medline: 33658529]
28. pcm-dpc / COVID-19. GitHub. URL: https://github.com/pcm-dpc/COVID-19 [accessed 2021-08-01]
29. Jajosky RA, Groseclose SL. Evaluation of reporting timeliness of public health surveillance systems for infectious diseases. BMC Public Health 2004 Jul 26;4:29 [FREE Full text] [doi: 10.1186/1471-2458-4-29] [Medline: 15274746]
30. Swaan C, van den Broek A, Kretzschmar M, Richardus JH. Timeliness of notification systems for infectious diseases: A systematic literature review. PLoS One 2018 Jun 14;13(6):e0198845 [FREE Full text] [doi: 10.1371/journal.pone.0198845] [Medline: 29902216]
31. Mavragani A. Tracking COVID-19 in Europe: Infodemiology Approach. JMIR Public Health Surveill 2020 Apr 20;6(2):e18941 [FREE Full text] [doi: 10.2196/18941] [Medline: 32250957]
32. 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]
33. 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]
34. Sousa-Pinto B, Anto A, Czarlewski W, Anto JM, Fonseca JA, Bousquet J. Assessment of the Impact of Media Coverage on COVID-19-Related Google Trends Data: Infodemiology Study. J Med Internet Res 2020 Aug 10;22(8):e19611 [FREE Full text] [doi: 10.2196/19611] [Medline: 32530816]
35. 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]
36. Schneider PP, van Gool CJ, Spreeuwenberg P, Hooiveld M, Donker GA, Barnett DJ, et al. Using web search queries to monitor influenza-like illness: an exploratory retrospective analysis, Netherlands, 2017/18 influenza season. Euro Surveill 2020 May;25(21):1 [FREE Full text] [doi: 10.2807/1560-7917.ES.2020.25.21.1900221] [Medline: 32489174]
37. 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]
38. Brownstein JS, Chu S, Marathe A, Marathe MV, Nguyen AT, Paolotti D, et al. Combining Participatory Influenza Surveillance with Modeling and Forecasting: Three Alternative Approaches. JMIR Public Health Surveill 2017 Nov 01;3(4):e83 [FREE Full text] [doi: 10.2196/publichealth.7344] [Medline: 29092812]

39. Kogan NE, Clemente L, Liautaud P, Kaashoek J, Link NB, Nguyen AT, et al. An early warning approach to monitor COVID-19 activity with multiple digital traces in near real time. Sci Adv 2021 Mar;7(10):1 [FREE Full text] [doi: 10.1126/sciadv.abd6989] [Medline: 33674304]
40. Altmann S, Milsom L, Zillessen H, Blasone R, Gerdon F, Bach R, et al. Acceptability of App-Based Contact Tracing for COVID-19: Cross-Country Survey Study. JMIR Mhealth Uhealth 2020 Aug 28;8(8):e19857 [FREE Full text] [doi: 10.2196/19857] [Medline: 32759102]
41. Immuni. URL: https://www.immuni.italia.it/ [accessed 2021-08-01]
42. 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]
43. Ramsetty A, Adams C. Impact of the digital divide in the age of COVID-19. J Am Med Inform Assoc 2020 Jul 01;27(7):1147-1148 [FREE Full text] [doi: 10.1093/jamia/ocaa078] [Medline: 32343813]

Abbreviations

CON: users accessing the system because of contact with a confirmed COVID-19 case
DSS: decision support system
DTW: dynamic time warping
SAX: Symbolic Aggregate approXimation
SYM: users accessing the system for the evaluation of symptoms

---

Edited by C Basch; submitted 16.04.21; peer-reviewed by Z Zhang, G Fico, A Rovetta; comments to author 01.07.21; revised version received 14.07.21; accepted 18.07.21; published 13.08.21.

Please cite as:
Tozzi AE, Gesualdo F, Urbani E, Sbenaglia A, Ascione R, Procopio N, Croci I, Rizzo C
Digital Surveillance Through an Online Decision Support Tool for COVID-19 Over One Year of the Pandemic in Italy: Observational Study
J Med Internet Res 2021;23(8):e29556
URL: https://www.jmir.org/2021/8/e29556
doi:10.2196/29556
PMID:34292866

©Alberto Eugenio Tozzi, Francesco Gesualdo, Emanuele Urbani, Alessandro Sbenaglia, Roberto Ascione, Nicola Procopio, Ileana Croci, Caterina Rizzo. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Changes in Public Response Associated With Various COVID-19 Restrictions in Ontario, Canada: Observational Infoveillance Study Using Social Media Time Series Data

Antony Chum[1,2*], PhD; Andrew Nielsen[1*], BPH; Zachary Bellows[1], BSc; Eddie Farrell[1], BSc; Pierre-Nicolas Durette[3], BSc; Juan M Banda[4], PhD; Gerald Cupchik[3], PhD

[1]Department of Applied Health Sciences, Brock University, St. Catharines, ON, Canada
[2]MAP Centre for Urban Health Solutions, Unity Health Toronto, Toronto, ON, Canada
[3]Department of Psychology, University of Toronto, Toronto, ON, Canada
[4]Department of Computer Science, College of Arts and Sciences, Georgia State University, Atlanta, GA, United States
*these authors contributed equally

Corresponding Author:
Antony Chum, PhD
Department of Applied Health Sciences
Brock University
1812 Sir Isaac Brock Way
St. Catharines, ON, L2S 3A1
Canada
Phone: 1 905 688 5550
Email: antony.chum@gmail.com

Abstract

Background: News media coverage of antimask protests, COVID-19 conspiracies, and pandemic politicization has overemphasized extreme views but has done little to represent views of the general public. Investigating the public's response to various pandemic restrictions can provide a more balanced assessment of current views, allowing policy makers to craft better public health messages in anticipation of poor reactions to controversial restrictions.

Objective: Using data from social media, this infoveillance study aims to understand the changes in public opinion associated with the implementation of COVID-19 restrictions (eg, business and school closures, regional lockdown differences, and additional public health restrictions, such as social distancing and masking).

Methods: COVID-19–related tweets in Ontario (n=1,150,362) were collected based on keywords between March 12 and October 31, 2020. Sentiment scores were calculated using the VADER (Valence Aware Dictionary and Sentiment Reasoner) algorithm for each tweet to represent its negative to positive emotion. Public health restrictions were identified using government and news media websites. Dynamic regression models with autoregressive integrated moving average errors were used to examine the association between public health restrictions and changes in public opinion over time (ie, collective attention, aggregate positive sentiment, and level of disagreement), controlling for the effects of confounders (ie, daily COVID-19 case counts, holidays, and COVID-19–related official updates).

Results: In addition to expected direct effects (eg, business closures led to decreased positive sentiment and increased disagreements), the impact of restrictions on public opinion was contextually driven. For example, the negative sentiment associated with business closures was reduced with higher COVID-19 case counts. While school closures and other restrictions (eg, masking, social distancing, and travel restrictions) generated increased collective attention, they did not have an effect on aggregate sentiment or the level of disagreement (ie, sentiment polarization). Partial (ie, region-targeted) lockdowns were associated with better public response (ie, higher number of tweets with net positive sentiment and lower levels of disagreement) compared to province-wide lockdowns.

Conclusions: Our study demonstrates the feasibility of a rapid and flexible method of evaluating the public response to pandemic restrictions using near real-time social media data. This information can help public health practitioners and policy makers anticipate public response to future pandemic restrictions and ensure adequate resources are dedicated to addressing increases in negative sentiment and levels of disagreement in the face of scientifically informed, but controversial, restrictions.

(J Med Internet Res 2021;23(8):e28716) doi:10.2196/28716

KEYWORDS

COVID-19; public opinion; social media; sentiment analysis; public health restrictions; infodemiology; infoveillance; coronavirus; evaluation

Introduction

Background

Since the identification of SARS-CoV-2 in late 2019 until February 7, 2021, there have been 106 million cases of COVID-19 infections worldwide along with 2.32 million deaths. To contain the spread of infection, many national and regional governments have implemented a series of public health restrictions, including travel restrictions, closing of nonessential businesses, school closures, mandatory masking, social distancing rules, and other restrictions on the movement of populations.

While news media coverage of the public response to these COVID-19 restrictions have highlighted the growing number of antimask protests, COVID-19 conspiracies, and pandemic politicization with extreme views, these characterizations may not necessarily represent the general public opinion and sentiment about pandemic restrictions. The objective of this study is to investigate the association between pandemic restrictions and COVID-19–related public sentiment (ie, collective attention, aggregate sentiment, and sentiment polarization and disagreement) using Twitter data. The development of novel methods to incorporate sentiment analysis into the evaluation of public health restrictions is important, since traditional methods of monitoring public reactions are often expensive and inefficient (eg, random representative surveys) and may suffer from limited coverage and significant delays.

Prior Relevant Studies

Previous research has emphasized that the use of social media by leaders and officials can lead to rapid dissemination of COVID-19–related information and influence of public policy [1]; however, the views of the general public, expressed via social media, should also be considered to inform effective pandemic response. Understanding how the public perceives these COVID-19 restrictions and information can inform public health messaging to maximize adherence to guidelines and reduce the spread of the virus. In a recent scoping review of studies related to COVID-19 and social media concerning the first outbreak from November 2019 to November 2020 [2], the authors noted a growing number of studies that document social media reaction to the COVID-19 pandemic to track and identify prevalent themes and concerns. While there is a larger body of literature that has identified changes in public opinions and perceptions over time using sentiment, topic, and content analysis of COVID-19–related social media content [3-8], the authors of the review noted that there is a scarcity of studies—at the time of publication in January 2021—that evaluate the impact of public health restrictions on public opinions (level of positive and negative emotions, level of disagreement, etc). However, some studies have begun to examine how COVID-19–related events (ie, COVID-19 case incidence, interventions, and news media) coincide with the frequency of COVID-19–related social media discussion (ie, collective attention).

In a study of COVID-19–related tweets from February 25 to March 30, 2020, in Belgium [9], researchers plotted tweet frequency alongside major COVID-19–related events, and they found that spikes in tweet frequency coincided with COVID-19 infections, stock market crashes, school closures, and infections of notable persons. The frequency of tweets about COVID-19–related topics has also been used to measure perceived susceptibility and severity of COVID-19 and was correlated with interventions, public events, and case counts [10]. Another descriptive study of COVID-19–related tweets from Australian states and territories detailed changes in aggregate sentiment trends in relation to COVID-19–related deaths and major COVID-19–related policy events [11] (eg, blocking arrivals from specific countries, expansion of testing criteria, and limits on outdoor gatherings). However, due to the lack of multivariate statistical modeling in the studies listed above, it was not possible to disentangle the independent contribution of these events on tweet frequency or aggregate sentiment and to investigate their relative importance in the shaping of public opinion.

Other studies have employed statistical models to understand factors that contribute to social media collective attention on COVID-19. In a study of the effects of COVID-19–related news coverage on collective attention [12]—measured by posts and comments on the r/coronavirus subreddit on reddit.com—between February 15 and May 15, 2020, researchers found, using linear regression, that the collective attention across the United Kingdom, the United States, Canada, and Italy was associated with daily COVID-19 incidence and COVID-19–related news articles. However, it is worth mentioning that the study did not include other factors that might also influence collective attention in their models, such as duration of business closure, the influence of holidays, and the introduction of restrictions including social distancing and mandatory face masks. The study focused mainly on collective attention (ie, comment and post frequency) but did not evaluate other indicators that might be more relevant to policy makers, such as the level of disagreement (eg, sentiment polarity) and aggregate sentiment (eg, positive to negative sentiment ratio) [4,5].

The limited number of studies that examined the association between COVID-19–related events, restrictions, and public opinion have typically approached the question from a descriptive manner, such as by graphically plotting major events and COVID-19 incidence on a timeline against COVID-19–related tweet frequency [9]. However, without considering the contribution of multiple factors simultaneously (business closures, school closures, holidays, other restrictions, etc), such as through the use of multivariate time series analysis, these studies may over- or understate the unique contribution of any given factor due to statistical confounding. To overcome this problem, our approach was informed by a prior study that used multivariate time series methods to analyze Twitter data, which accounted for multiple control variables, serial autocorrelation, and seasonal fluctuations and trends [13].

Additionally, the previous studies were unable to quantify the strength of the relationships between exposure (eg, days of business closure) and relevant public opinion outcomes (eg, level of negative sentiment). Our study will bridge this gap in the literature by using a dynamic regression approach to understand the unique contribution of restriction specifications (ie, business closures, school closures, announcements of masking and social distancing measures, and regional lockdown differences) on public opinion, while also taking into account the influence of contextual factors, including case counts, holidays, and COVID-19–related official updates. Our research question is as follows: What is the association between COVID-19 public health restrictions and measures of public opinion (ie, collective attention, positive to negative sentiment ratio, and level of disagreement) while accounting for potential confounding factors?

Methods

Twitter Data Collection

Data from our study were drawn from the largest COVID-19–related Twitter data set [14]. It was constructed using the following data-driven selection of keywords: COVD19, CoronavirusPandemic, COVID-19, 2019nCoV, CoronaOutbreak, coronavirus, WuhanVirus, covid19, coronaviruspandemic, covid-19, 2019ncov, and coronaoutbreak. The Social Media Mining Toolkit [15] was used to collect all tweets worldwide with the keywords mentioned above starting on March 12, 2020. Further details about the data collection process can be found in a previous paper [14]. We used the cleaned data set of English-only tweets, with retweets filtered out. A retweet is the sharing of a tweet without any added comments; however, quoted tweets (ie, sharing a previous tweet along with one's own comment) were included in the data.

To identify a subset of tweets originating from Ontario, Canada, geographic coordinates were used for tweets with geolocation enabled. For tweets that did not have geolocation enabled, our team created an algorithm that matched the text of the user-defined location to a standard gazetteer at GeoNames [16]. The gazetteer data contain alternative spellings for cities across different languages and include various airport codes used for matching (eg, YTO and YYZ for Toronto). We used a list of locations that had a population of 1000 or greater. When inferring location based on user input, our algorithm matches to a city with a unique name. For cities that share the same name with other cities, the algorithm attempts to find a match based on country and/or state identifiers in the text. If there is no state or country data in the text (eg, "London" only), the tweet is matched to the place with the highest population; in this case it would be London, England, UK. Matching to the largest population center, in cases where no further information is available, was based on the assumption that people from the largest cities are more likely to leave out further country or regional identifiers, while those in smaller cities that share the same name with larger cities are more likely to include further regional information. If no match is made at the city or town level, the text is then matched to a higher-level geographical unit (ie, state, region, or province) and then to a country. Out of the subset of all tweets with user-entered location text, our program matched 89.9% to a GeoName ID. A link to our GitHub repository for the algorithm is available [17]. Our program also examined any Unicode data (eg, a flag emoji) entered by users in lieu of country-level information. We randomly sampled 250 matches to ensure that the matches were made according to the algorithm described above. In total, we identified 2,649,317 tweets originating from Canada between March 12 and October 31, 2020, 43.4% of which (1,150,362 tweets) were from Ontario.

Sentiment Analysis

Once we collected the COVID-19 Twitter data, we conducted sentiment analysis using the VADER (Valence Aware Dictionary and Sentiment Reasoner) algorithm, which assigned a sentiment score (–1 to +1) to each tweet that represents a polarity—negative or positive—and a strength of emotion for the tweets. Table 1 presents examples of positive, neutral, and negative tweets and their VADER-assigned sentiment scores. In a prior study [18], scoring by the program had an r=0.88 correlation with gold-standard ground truth (ie, the mean sentiment rating from 20 prescreened and appropriately trained human raters). Scores of –0.05 and under were negative, scores of +0.05 and above were positive, and scores in between were neutral. These thresholds are conventional for studies using VADER [19,20], and classification by human raters was found to be well-matched to VADER results when using these scoring thresholds [21].

Table 1. Examples of positive, neutral, and negative tweets with VADER[a]-assigned sentiment scores.

Sentiment score	Classification	Tweet
0.93	Positive	“Thank you so much @johnkrasinski for this series! I think it helped remind everyone how much good there is in the world. I really hope the silver lining of COVID-19 is people continue to be kinder to one another and truly realize we're all in this together.”
0.65	Positive	“@celliottability notes that Ontario has made great strides on COVID-19 testing and contact tracing. Anyone who wants to get a COVID-19 test can do so, even if they don’t have symptoms”
0.03	Neutral	“#SSHRCResearchers Helen Kennedy and Sarah Atkinson look at how the industry is adapting to the new reality of #COVID19”
–0.04	Neutral	“Why you should wear a #mask #COVID10 @ottawahealth”
–0.40	Negative	“COVID-19 Compliance: One-in-five Canadians making little to no effort to stop coronavirus spread”
–0.57	Negative	“Because the Chinese just hate witchcraft. Riiiiight... Cough, feng shui, cough #WuhanVirus #COVID19”

[a] VADER: Valence Aware Dictionary and Sentiment Reasoner.

Study Outcomes

Overview

To study public opinion on COVID-19–related public restrictions using Twitter data in a comprehensive manner, we considered (1) the collective attention on COVID-19 measured by the level of COVID-19–related discussion (ie, COVID-19–related tweet frequency); (2) the aggregated sentiment level, measured using a positive to negative sentiment ratio; and (3) the level of disagreement, or sentiment polarity, measured by the Gini index.

COVID-19–Related Discussion: Tweet Frequency

We used tweet frequency to represent the level of participation in COVID-19–related discussion on Twitter on a specific day. Prior studies have utilized social media activity data (ie, Twitter and Weibo post frequency) to identify collective attention with regard to COVID-19 interventions and events [5,12]. We have included tweet frequency to estimate how public health restriction can influence COVID-19–related collective attention, which may provide a useful metric that policy makers can use to identify potential areas of concern at the population level.

Aggregate Sentiment

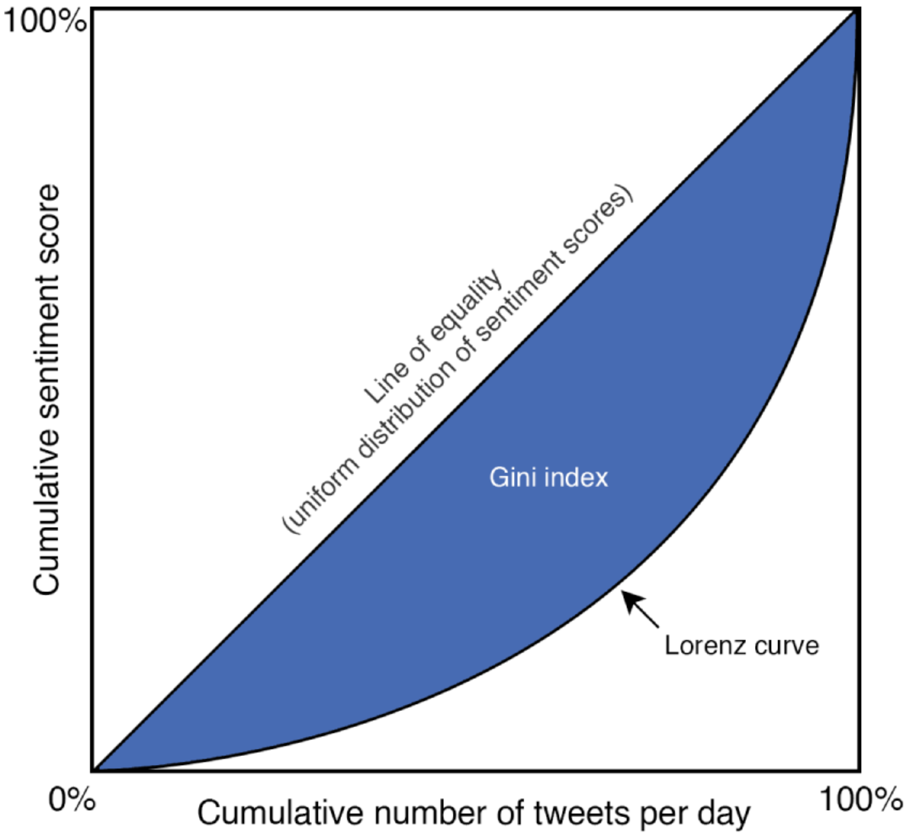
To determine the aggregate sentiment of a particular day, a value was derived for each day that represents the ratio of positive to negative sentiment, expressed by the following:

where $M_{t,pos}$ is the total count of positive tweets with sentiment scores greater than 0.05, and $M_{t,neg}$ is the count of negative tweets with sentiment scores lower than –0.05. The natural log transformation is used to avoid excessively large ratios. This specific formula for sentiment aggregation to measure the net positive sentiment has been commonly used in prior literature of Twitter sentiment analysis, since it takes into account the number of Twitter users on a given day [22,23]. The use of positive to negative sentiment ratios have been predictive of group-level behaviors, such as stock market [23] and movie-going behaviors [24,25]. These previous studies excluded neutral tweets, since they tend to represent objective or informational statements, often coming from nonprofits, governments, or academic institutions. We did not include them in our measure, since we aimed to measure the subjective perspectives and views of individuals characterized by negative and positive emotions. In addition, a public health restriction that is associated with significantly more positive than negative tweets has a greater chance of being accepted and may reflect a higher level of public compliance.

Gini Index

A Gini index was derived to measure the level of disagreement, or sentiment polarization, in COVID-19–related tweets. Although the Gini index is typically used in the literature to describe income inequality, this index has been used to measure inequality in other areas of social interest, such as opportunity for social mobility [26], educational attainment [27], public transit availability [28], and movie preferences [29]. A Gini index of zero represents the lowest level of disagreement (ie, perfect equality of scores), and a higher Gini index represents greater differences in the sentiment scores across tweets on a particular day. For example, a Gini index of 0.30 means that 30% of the sentiment scores would have to be redistributed in order for everyone’s score to be the same. The Gini index is calculated based on the area between (1) the line of equality and (2) the Lorenz curve, as shown in Figure 1.

Figure 1. Graphical representation of using the Gini index to measure sentiment disparity.


100%
Cumulative sentiment score
Line of equality
(uniform distribution of sentiment scores)
Gini index
Lorenz curve
0%
Cumulative number of tweets per day
100%

The 45° line of equality, where x=y, is the hypothetical situation of uniform distribution where each tweet, in the same day, exhibits the same sentiment score. By plotting the cumulative sentiment score on a given day against the cumulative number of tweets, the Lorenz curve can be used to characterize sentiment score disparity (ie, visually represent how a range of tweets, from those with the lowest to highest sentiment scores, contribute to the relative increases in the cumulative score, where a more concave Lorenz curve represents greater disparity). To create a daily Lorenz curve, we started by rescaling each tweet-level sentiment score (–1 to +1) to a range from 1 to 100, because the standard calculations cannot include negative values. We then ordered tweets from the lowest to the highest sentiment scores and plotted the cumulative number of tweets against the cumulative tweet sentiment score. Next, the Gini index was calculated by finding the area under the line of equality and above the Lorenz curve (eg, shaded blue area in Figure 1). This method of calculating a Lorenz curve and Gini index was repeated for each day in our data set. Our Gini index (G) is calculated using the following equation:

where X_k is the cumulative proportion of tweets over n number of tweets in a given day (from k=0,...n,), and Y_k is the cumulative proportion of sentiments in a given day.

Creating the Ontario COVID-19 Timeline

Overview

We created a comprehensive timeline of COVID-19–related restrictions and events in Ontario by consulting with the COVID-19 intervention timeline created by the Canadian Institute for Health Information [30], the timeline of COVID-19 events created by Public Health Ontario [31], and timelines that were created by news media [32,33]. This full timeline used for the study is available in Multimedia Appendix 1. In Ontario, key events on the timeline include the following:

1. The declaration of a state of emergency on March 17, 2020, which led to the closing of all nonessential businesses and schools.
2. The closure of the US-Canada border to nonessential travelers on March 21, 2020.
3. The partial reopening of selected regions in Ontario that began on June 12, 2020.
4. The reopening of nearly all businesses and public places across Ontario, with restrictions, by August 12, 2020.
5. The restrictions to reduced private gatherings that were reinstated on September 19, 2020.
6. The restrictions on restaurants, bars, banquet halls, and gyms that were reinstated on October 3, 2020, in selected urban regions.

For the purpose of our study, we focused on four dimensions of the public health restrictions, including (1) business closures, (2) school closures, (3) regional lockdown differences (ie, partial vs province-wide lockdown), and (4) additional public health measures (eg, travel, social distancing, and masking).

Business Closures

Ontario implemented closures and limitations on nonessential businesses to help control the spread of COVID-19. With the exception of essential businesses—including stores that sold food, big box retailers, pharmacies, and alcohol stores—that stayed open, many businesses were closed or offered limited services (eg, restaurants were limited to providing delivery or take-out services only). Given the importance of business and retail services to Ontario residents, we considered the cumulative effect of business closures. An urban-centric approach was used to define business closure, since the majority of Ontarians live in major urban centers (71.7% as of 2019) [34]. While rural areas reopened earlier in the summer of 2020, for the purposes of the timeline, we did not consider businesses across the province to be reopened until it was the case in all major population centers, with some limits on capacities. To account for the influence of the earlier rural reopening, we included an adjustment variable to indicate the partial reopening of Ontario. To construct the business closure variable, we first created a binary variable to indicate, for each day on the timeline, whether nonessential businesses were closed due to restrictions. For each consecutive day of closure, we created a cumulative variable to consider effects associated with the duration of closure (eg, 1 for the first day of closure and 10 for the 10th day of closure). Additionally, we hypothesized that each additional day of closure had an additive but diminishing effect (ie, logarithmic growth) because each additional day of the closure could have a normalizing effect due to adaptation; therefore, we derived the natural log cumulative business closure variable to be used in our regression models.

School Closures

We considered primary and secondary school closures to be a significant restriction that impacts a large number of Ontario families. Moreover, the closure of primary and secondary schools would lead to the need for parents to make accommodations to provide childcare. Days for school closure due to COVID-19 restrictions were represented through a binary variable. Universities and colleges were not considered, as students are older and able to care for themselves, therefore causing less disruption. We hypothesized a logarithmic growth effect on the experience of school closure because each additional day of the closure could have a normalizing effect, where each additional day of closure has an additive but diminishing effect, due to adaptation and adjustment to new childcare arrangements and work accommodations.

Regional Lockdown Differences

Over the course of the study period, Ontario implemented province-wide lockdowns and partial lockdowns, where the latter focused on dense urban areas (eg, Toronto, Peel, and Ottawa) to implement a targeted approach to pandemic restrictions. We categorized days in our time series into three groups: (1) province-wide lockdown, (2) partial lockdown, and (3) no lockdown. We expected these variations in lockdown conditions to have an effect on social media discussion and sentiment. Decisions around the implementation of partial versus province-wide lockdowns were controversial [35], with diverging beliefs around the benefits of a province-wide lockdown (eg, under partial lockdown, some people may travel to an adjacent region with no lockdown to visit the gym) and the benefits of partial lockdowns (eg, partial lockdown allows for a flexible approach tailored to local COVID-19 infection rates to minimize economic impacts).

Additional Public Health Measures

Between March 12 and October 31, 2020, there were a number of additional restrictions put in place by the Ontario and federal governments to reduce the spread of COVID-19. These include measures such as nonessential travel restrictions (eg, US-Canada border), mandatory quarantine for travelers, limits on indoor and outdoor gatherings, health and safety bylaws for businesses such as sanitizer and plexiglass, and social distancing and mask policies for the general population. Given the overlapping nature of multiple public health measures, which often target specific concerns (ie, travel, distancing, and hygiene), we only considered the day a restriction was announced. Unlike business and school closures, these additional public health measures remained enforced for the duration of our data set. We characterized each day as having either (1) a new or updated restriction announced or (2) no restrictions announced.

Control Variables

In order to adjust for other contextual factors that may also influence COVID-19–related public opinion, we included the following variables as control factors: (1) COVID-19–related official updates, (2) statutory holidays, (3) COVID-19 daily incidence for Ontario, and (4) COVID-19 daily incidence for Canada, excluding Ontario.

COVID-19–Related Official Updates

Multiple official COVID-19–development announcements have been released over the course of the pandemic, including press conferences for major events (ie, case counts and mortality milestones), new screening guidelines, and provincial reopening plans, as well as notable COVID-19 developments (eg, new evidence on the effectiveness of nonmedical masks) from the World Health Organization, Ontario Hospital Administration, and government officials. While additional public health measures detail restrictions enforced on the population, COVID-19–related official updates are meant only to provide useful information about COVID-19 events. For example, after the announcement that Canada had surpassed 100,000 COVID-19 cases, we might expect that people would take to social media to express their emotions about this information.

Statutory Holidays

There is prior evidence that public sentiment and frequency of posts on holidays systematically differ from those on nonholidays [36]. There were also public concerns that travel and social gathering plans over holidays, and long weekends, may promote COVID-19 infections [37] and, in turn, public sentiment concerning COVID-19. Therefore, we included

Canadian statutory holidays in our models as an adjustment variable. In our study period, seven holidays in 2020 were identified: Good Friday (April 10), Easter (April 12), Victoria Day (May 18), Canada Day (July 1), Civic Holiday (August 3), Labour Day (September 7), and Thanksgiving (October 12). If the holiday was part of a long weekend, the entire weekend was coded as a holiday. For example, Labour Day was on Monday, September 7; therefore, Saturday, September 5 and Sunday, September 6 were also coded as a holiday.

COVID-19 Daily Incidence

COVID-19 new daily case counts at the provincial and national levels were a major focus in news media and a significant factor that could influence public opinion and collective attention on COVID-19. Case information is based on the Public Health Case and Contact Management Solution [31], which is Ontario's primary disease-reporting system. Case counts for Canada were drawn from the COVID-19 Data Repository at Johns Hopkins University [38]. We subtracted the Ontario case counts from the Canada case counts so the national numbers were deduplicated.

Statistical Analysis

Our study combined an autoregressive integrated moving average (ARIMA) approach to time series modeling with regression methods in order to examine the associations between public health restrictions and changes in sentiment measures over time [39]. These are generally known as dynamic regression models, and they are typically used to generate forecasts [40] but are also useful for the purpose of explanatory modeling (ie, understanding the relationship between multiple time series variables), as is the purpose of our study. They take on the form where the outcome time series y_t is modeled as a function of k explanatory variables $(x_{1,t},...x_{k,t})$, where n_t is allowed to be autocorrelated, using ARIMA errors:

The ARIMA error n_t may contain (1) autoregressive (AR) terms used to determine the relationship between the current observation and previous observations, (2) moving average (MA) terms to determine the relationship between current observation and previous error, and (3) differencing terms to stationarize the time series outcome if necessary. For parameter estimations of ARIMA error terms (ie, the content of n_t), the auto.arima() function was used in the *forecast* package in R, version 4.0.3 (The R Foundation). The purpose of using this function is to fit the most appropriate ARIMA model according to Akaike information criterion (AIC) values; the AIC quantifies the model's goodness of fit (ie, lower is a better fit). The function searches across a number of candidate models, selects the appropriate number of AR and MA terms based on minimization of the AIC, and applies the appropriate number of differencing terms to stationarize the outcome time series [41]. The selection of terms is denoted using (*p*, *d*, *q*), where p is the number of AR terms, d is the degree of differencing, and q is the number of MA terms. For example, if auto.arima() determined that ARIMA (1, 0, 0) was most appropriate for the dynamic regression model, this would indicate that a model using one AR term produces the greatest minimization of the AIC.

We constructed three models, with Model 1 for the frequency of tweets concerning COVID-19 each day (ie, collective attention), Model 2 for the aggregate sentiment score representing the ratio of positive to negative tweets each day, and Model 3 for the level of sentiment disparity within each day, using the Gini index. The outcomes were deseasonalized using the ts() function in R, since there is a tendency for more Twitter activities on weekdays over weekends. Following the deseasonalizing procedure, we used the augmented Dickey-Fuller and Kwiatkowski-Phillips-Schmidt-Shin tests for stationarity. If the trend was not stationary, differencing for stationarity would be handled by the auto.arima() function. Each outcome was regressed on all seven predictors mentioned above, and regressors that were not significantly associated with the outcome (at *P*<.05) were subsequently removed. All statistical analyses were completed on RStudio Cloud (updated to January 20, 2021) using R, version 4.0.3.

Results

Overview

We collected 1,150,362 COVID-19–related tweets that originated from Ontario, Canada, between the period of March 12 and October 31, 2020, which consisted of 235 days. The mean daily tweet frequency was 4933 (SD 1065). The descriptive statistics and bivariate associations for outcomes and regressors are presented in Table 2. Kruskal-Wallis one-way analysis of variance tests were conducted to identify significant differences in mean tweet frequency, aggregate sentiment, and Gini index across levels of the regressors. The mean Gini index was 24.19 (SD 0.85), meaning that, on average, 24.18% of the scores would have to be redistributed for every tweet to have the same level of sentiment. The aggregate positive to negative tweet ratio was 34.57 (SD 7.92), meaning that more tweets were considered positive than negative based on the sentiment analysis. The univariate time series for frequency of tweets, the aggregate sentiment score, and the Gini index are displayed in Figure 2.

After deseasonalizing the three outcome variables, augmented Dickey-Fuller and Kwiatkowski-Phillips-Schmidt-Shin tests returned large *P* values (*P*>.10) across all three variables, which provides evidence that they were stationary. This is further confirmed through visual inspection, and the fact that the auto.arima() did not require the inclusion of differencing terms in any subsequent models. An autocorrelation function plot of each outcome (Figure 3) shows no significant autocorrelations, indicating that the residuals behave like white noise and, therefore, do not exhibit temporal autocorrelations.

Table 2. Descriptive statistics and bivariate associations for outcomes and regressors.

Outcomes and regressors	Days with event, March 12 to October 31 (n=235), n (%)	Tweet frequency (days with condition)		Gini index (days with condition)		Positive to negative ratio (days with condition)	
		Mean (SD)	*P* value[a]	Mean (SD)	*P* value[a]	Mean (SD)	*P* value[a]
Business closure							
Nonessential businesses closed	143 (60.9)	5384.90 (1136.55)	<.001	24.33 (0.78)	<.001	35.28 (10.22)	.25
Nonessential businesses open	92 (39.1)	4127.85 (1285.98)		23.95 (0.92)		33.46 (10.98)	
School closure							
Schools open	126 (53.6)	4302.24 (1222.53)	<.001	24.03 (0.90)	.003	33.27 (10.84)	.045
Schools closed due to COVID-19	109 (46.4)	5575.42 (1141.91)		24.36 (0.76)		36.06 (10.03)	
Additional restrictions							
No restriction announcements	223 (94.9)	4876.51 (1354.63)	.45	24.22 (0.84)	.003	34.23 (10.62)	.04
New or updated restriction announced	12 (5.1)	5195.16 (1122.22)		23.46 (0.75)		40.83 (6.40)	
Regional differences in lockdown							
Province-wide lockdown	169 (71.9)	5137.50 (1334.11)	<.001	24.23 (0.85)	.17	34.44 (9.85)	.08
Partial lockdown	61 (23.0)	4347.68 (1094.77)		24.08 (0.85)		35.66 (12.07)	
No regions under lockdown	5 (2.1)	3271.60 (1587.16)		23.76 (0.93)		25.34 (10.76)	
Statutory holidays							
Holidays (with attached weekends)	17 (7.2)	4016.06 (1249.85)	.005	24.86 (0.52)	<.001	24.41 (9.34)	<.001
Nonholidays	218 (92.8)	4961.15 (1329.05)		24.13 (0.85)		35.36 (10.23)	
New COVID-19 case counts (in hundreds of cases)							
Low (0-1.57)	78 (33.2)	4158.34 (963.00)	<.001	24.05 (0.73)	.18	35.01 (10.15)	.40
Medium (1.58-4.04)	76 (32.3)	5275.28 (1282.49)		24.21 (0.89)		35.51 (10.21)	
High (4.05+)	81 (34.5)	5241.12 (1434.39)		24.28 (0.92)		33.25 (11.20)	
New COVID-19 case counts in Canada (in hundreds of cases)							
Low (0-4.53)	78 (33.2)	4185.12 (1143.35)	<.001	24.22 (0.75)	.17	33.70 (10.67)	.24
Medium (4.54-12.36)	77 (32.8)	5174.81 (1211.07)		24.08 (0.95)		36.06 (10.23)	
High (12.37+)	80 (34.0)	5311.31 (1382.81)		24.26 (0.86)		33.99 (10.70)	
Official announcements of COVID-19 developments (WHO[b] declarations, release of reopening plans, etc)							
No announcement	218 (92.8)	4848.34 (1342.49)	.10	24.22 (0.01)	.01	34.35 (10.31)	.21
Announcement	17 (7.2)	5462.65 (1258.50)		23.69 (0.92)		37.34 (13.27)	

[a] *P* values were calculated for Kruskal-Wallis tests for differences in means across levels.

[b] WHO: World Health Organization.

Figure 2. Daily tweet frequency, aggregate sentiment, and Gini index time series.

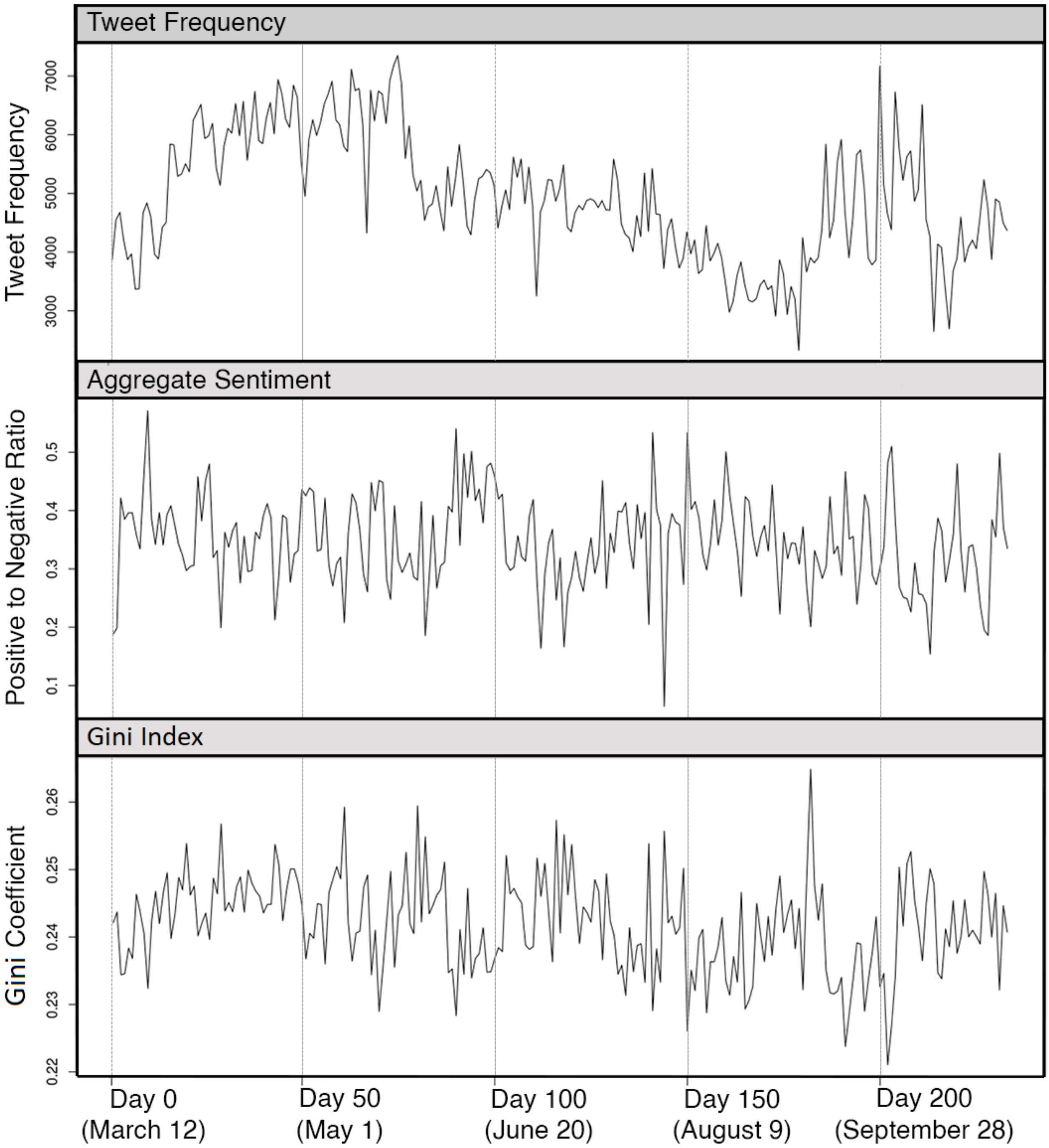
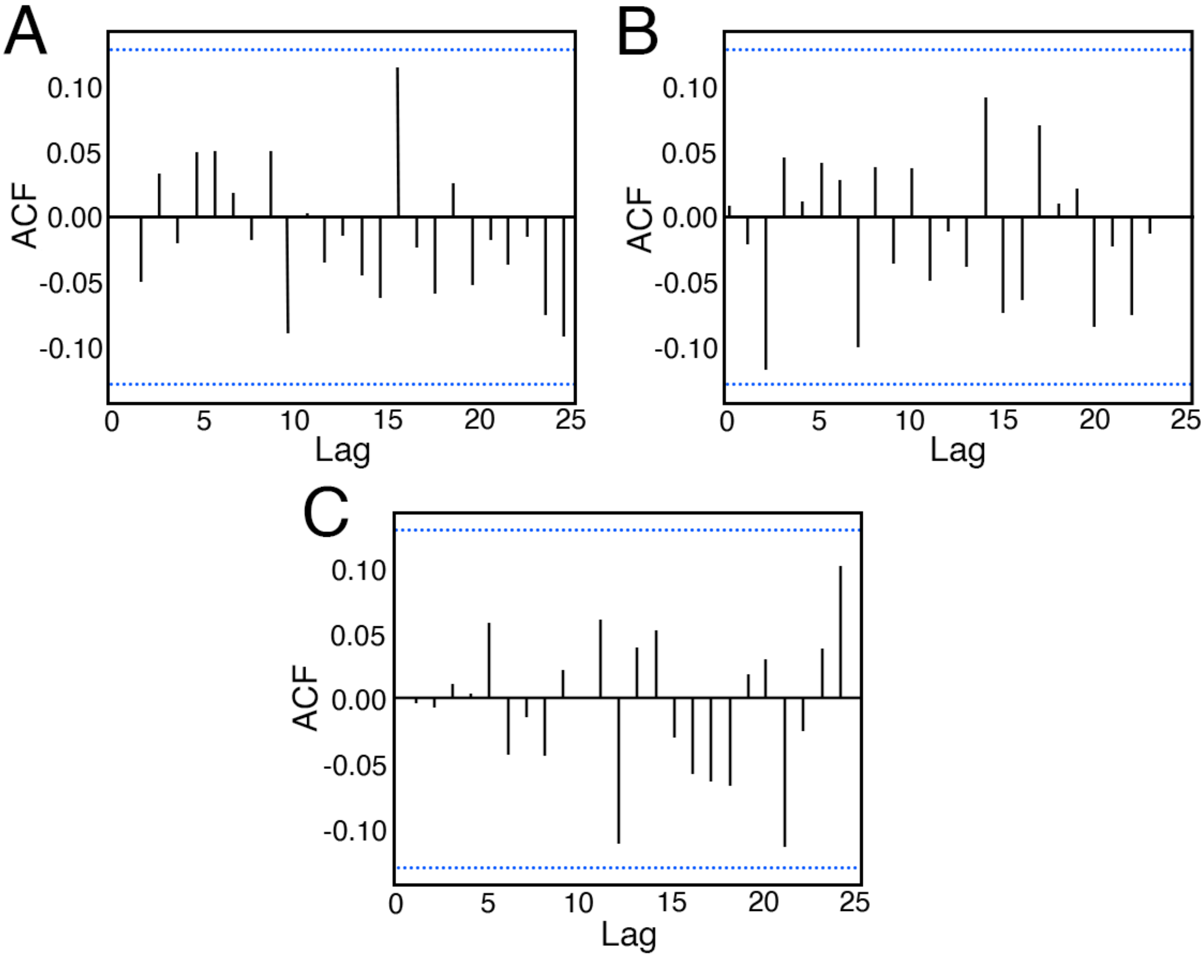


Figure 3. Autocorrelation function (ACF) plots for (A) COVID-19–related tweet frequency, (B) aggregate sentiment, and (C) Gini index.



COVID-19–Related Tweet Frequency

Auto.arima() selected two AR terms (2, 0, 0) to be used in the ARIMA model predicting COVID-19–related tweet frequency (Table 3). Inclusion of predictor terms in the model improved the value of the AIC by 5%. Additional days of business and school closures were associated with more tweets in a nonlinear manner, where one additional day of closure had a stronger effect in the earlier part of the closure compared to the later parts. In other words, the effect of closure had a diminishing effect on tweet frequency with each additional day of closure. Each 10% increase in the duration of business closure (ie, 196 × log_e [1.1] = 18.6) was associated with an increase of 18.6 tweets (95% CI 11.5-25.8). Each 10% increase in the duration of school closure (ie, 130 × log_e [1.1] = 12.3) was associated with an increase of 12.3 tweets (95% CI 5.7-18.9). Figure 4 plots the rate of increase of tweet frequency associated with business and school closures. The announcement of additional public health restrictions was associated with 544 additional tweets (95% CI 178-910). Based on the statistically significant interaction between new daily COVID-19 cases in Ontario and lockdown condition (ie, Ontario case counts by province-wide vs partial lockdown; *P*<.001), new COVID-19 cases had a different effect on tweet frequency depending on the lockdown condition. Under province-wide lockdown, each increment of 100 new COVID-19 cases was associated with 391 additional tweets, and under partial lockdown, each increment of 100 new cases was associated with 134 additional tweets. The effect of new COVID-19 cases under the no lockdown condition was not different compared to the province-wide lockdown condition (*P*=.49). Compared to nonholidays, statutory holidays and their connected weekends saw a decrease of 385 tweets (95% CI –761 to –7.8). Each additional increment of 100 new cases across Canada, excluding Ontario, was associated with 46.2 additional tweets (95% CI 20.9-71.6). Days with an official COVID-19–related update saw an additional 373 tweets (95% CI 95.4-650) compared to days without any updates.

Table 3. Model 1: dynamic regression model predicting daily tweet frequency with ARIMA[a] error term (2, 0, 0).

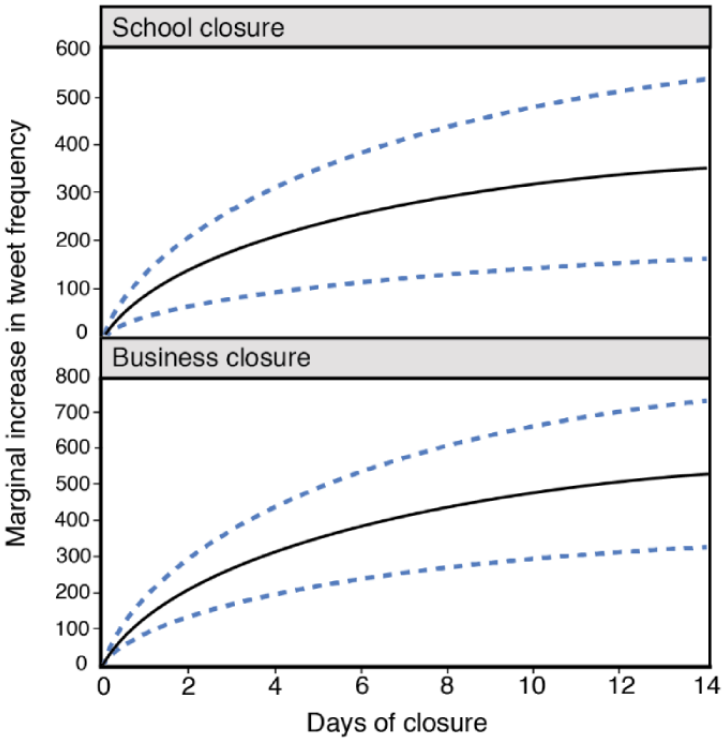
Measures	Tweet frequency	*P* value
Predictors of daily tweet frequency, estimate of effect (95% CI)		
Intercept	3143 (2837 to 3450)	<.001
Statutory holidays (1 for holidays, 0 for nonholidays)	–385 (–761 to –7.8)	.04
Business closure (increase in 1 log day)	196 (121 to 271)	<.001
School closure	130 (60.1 to 199)	<.001
Additional measures	544 (178 to 910)	.003
New COVID-19 case counts (in hundreds of cases)	391 (311 to 470)	<.001
New COVID-19 case counts in Canada, excluding Ontario (in hundreds of cases)	46.20 (20.9 to 71.6)	<.001
Official COVID-19–related updates	373 (95.4 to 650)	.008
Regional differences in lockdown		
Province-wide lockdown: regions are in the same stage of lockdown	Reference group	N/A[b]
Partial lockdown: regions are in different stages of lockdown	140 (–343 to 624)	.53
No lockdown: regions are not under lockdown	–440 (–1513 to 632)	.43
Regions are in different stages of lockdown × new cases	–257 (–361 to –153)	<.001
Regions are not in lockdown × new cases	1219 (–2161 to 4599)	.49
Goodness of fit		
With covariates, AIC[c]	3693.89	N/A
Without covariates, AIC	3873.20	N/A

[a] ARIMA: autoregressive integrated moving average.

[b] N/A: not applicable.

[c] AIC: Akaike information criterion.

Figure 4. Estimated marginal increases in tweet frequency associated with increases in number of days of business and school closures, holding all other factors constant. The black line represents the estimated change in positive to negative sentiment ratio, and the dotted blue lines represent the 95% confidence interval.



Positive to Negative Tweet Sentiment Ratio

One AR term was used in the (1, 0, 0) ARIMA model predicting positive to negative tweet sentiment ratio (Table 4). Compared to the empty model, the inclusion of predictor variables improved the model AIC by 6.5%. While higher COVID-19 case counts in Ontario had the effect of reducing the positive to negative ratio of tweet sentiment, where each increment of 100 new cases was associated with –0.98 in the aggregate sentiment ratio (95% –1.81 to –0.16) during the period where no business closures were in effect, the impact of new Ontario COVID-19 cases on the sentiment ratio changed once business closures were introduced, as indicated by the significant interaction term in Table 4 (ie, business closed × Ontario new cases; *P*=.02). To facilitate interpretation of the three-way nonlinear relationship, we plotted the change in predicted aggregate sentiment ratio from day 0 to day 10 of a business closure period given four case-count scenarios, where case counts were held constant at 50, 100, 150, and 200 over the closure period, as shown in Figure 5. In short, given everything else being equal, while higher case counts reduced sentiment ratio in a direct manner, higher case counts also reduced the negative effect associated with an additional day of business closure. Compared to days when Ontario was in a province-wide lockdown, a partial lockdown was associated with an increase in sentiment ratio of 5.75. The sentiment ratio was lower on statutory holidays compared to nonholidays (–6.22, 95% CI –10.3 to –2.12). New COVID-19 cases across Canada, excluding Ontario, were not associated with a change in sentiment ratio.

Table 4. Model 2: dynamic regression model predicting positive to negative ratio with ARIMA[a] error term (1, 0, 0).

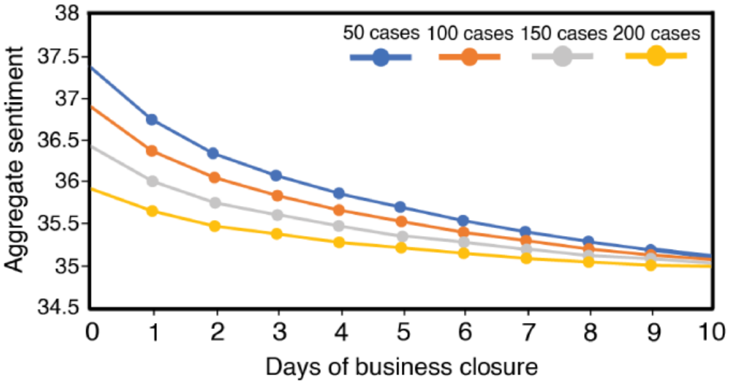
Measures	Positive to negative ratio	*P* value
Predictors of positive to negative ratio, estimate of effect (95% CI)		
Intercept	37.90 (34.60 to 41.20)	<.001
Statutory holidays (1 for holidays, 0 for nonholidays)	–6.22 (–10.30 to –2.12)	.002
Business closure (log transformed)	–1.14 (–2.26 to –0.01)	.046
Regional differences in lockdown		
Regions are in the same stage of lockdown	Reference group	N/A[b]
Regions are in different stages of lockdown	5.75 (2.16 to 9.33)	.001
Regions are not under lockdown	–10.50 (–18.70 to –2.29)	.01
New COVID-19 case counts in Canada, excluding Ontario (in hundreds of cases)	0.17 (–0.14 to 0.48)	.29
New COVID-19 case counts (in hundreds of cases)	–0.98 (–1.81 to –0.16)	.02
Business closed × new cases (increase of 1 log unit in business closure + 100 new cases)	0.37 (0.04 to 0.70)	.02
Goodness of fit		
With covariates, AIC[c]	1612.43	N/A
Without covariates, AIC	1723.89	N/A

[a] ARIMA: autoregressive integrated moving average.

[b] N/A: not applicable.

[c] AIC: Akaike information criterion.

Figure 5. Predicted change in positive to negative sentiment ratio from day 0 to day 10 of a business closure period, varying by new COVID-19 case counts in Ontario (holding all other factors constant).



Sentiment Disparity Measured by the Gini Index

One AR and two MA terms (1, 0, 2) were used in the ARIMA model predicting sentiment disparity (Table 5). Compared to the empty model (ie, with no predictors), the inclusion of predictors improved the model's AIC value by 20%. A higher Gini index represents a more polarized range of sentiments across COVID-19–related tweets in a given day (ie, more disparities in the sentiment scores). Each 10% increase in the duration of business closure (ie, $0.113 \times log_e [1.1] = 0.0107$) was associated with an increased Gini index of 0.01 (95% CI 0.005-0.01). Compared to days with a province-wide lockdown, days with a partial lockdown were associated with a 0.738 reduction in the Gini index (95% CI –1.19 to –0.283). We also found evidence that lockdown conditions can modify the effect of new COVID-19 case counts in Ontario on the Gini index, where each increment of 100 new cases was associated with a decrease of 0.11 in the Gini index (95% CI 0.01-0.21) while Ontario was under partial lockdown, but the Gini index remained unchanged with additional COVID-19 cases under province-wide lockdown (0.00, 95% CI –0.07 to 0.07). The Gini index was higher on statutory holidays compared to nonholidays (0.44, 95% CI 0.08-0.81). New COVID-19 cases in Canada, excluding Ontario, were not associated with the Gini index (95% CI –0.04 to 0.01).

Table 5. Model 3: dynamic regression model predicting the Gini index with ARIMA[a] error term (1, 0, 2).

Measures	Gini index	*P* value
Predictors of the Gini index, estimate of effect (95% CI)		
Intercept	23.90 (23.60 to 24.20)	<.001
Statutory holidays (1 for holidays, 0 for nonholidays)	0.44 (0.08 to 0.81)	.02
Business closure	0.11 (0.05 to 0.17)	<.001
Regional differences in lockdown		
Regions are in the same stage of lockdown	Reference group	N/A[b]
Regions are in different stages of lockdown	–0.738 (–1.19 to –0.28)	.001
Regions are not under lockdown	0.16 (–0.89 to 1.22)	.77
New COVID-19 case counts in Canada, excluding Ontario (in hundreds of cases)	–0.01 (–0.04 to 0.01)	.31
New COVID-19 case counts (in hundreds of cases)	0.00 (–0.07 to 0.07)	.99
Regions are in different stages of lockdown × new cases	0.11 (0.01 to 0.21)	.02
Regions are not under lockdown × new cases	–1.98 (–5.06 to 1.10)	.21
Goodness of fit		
With covariates, AIC[c]	461.82	N/A
Without covariates, AIC	573.98	N/A

[a] ARIMA: autoregressive integrated moving average.

[b] N/A: not applicable.

[c] AIC: Akaike information criterion.

Discussion

Principal Findings

Our study found significant associations between COVID-19 restrictions and public opinion. In summary, additional days of business closures were associated with collective attention (ie, COVID-19–related tweet frequency) and increased levels of disagreement (ie, sentiment polarity). While business closures reduced aggregate sentiment (ie, the net number of tweets with positive sentiment), additional COVID-19 cases reduced the impact of business closures on overall sentiment. In other words, the model shows that people were more accepting of additional business closure days if the cases were high. While additional days of school closures were associated with collective attention, with diminishing effects for each additional day, school closures were not associated with aggregate sentiment or levels of disagreement.

Compared to province-wide lockdowns, partial lockdowns were associated with increased aggregate sentiment (ie, net number of tweets with positive sentiment) and decreased levels of disagreement. Partial lockdowns, compared to province-wide lockdowns, were associated with decreased collective attention; they also reduced the effect of additional COVID-19 case counts on collective attention. In other words, while new COVID-19 case counts increased collective attention, this effect was reduced under partial lockdown compared to province-wide lockdown. Finally, we found that the announcement of other restrictions (eg, social distancing, masking, and travel restrictions) led to increased collective attention but were not associated with changes in aggregate sentiment or level of disagreement.

Comparison With Prior Literature

While our study was focused on investigating the unique impact of multiple pandemic restrictions on changes in public opinion over time, which has not been examined in prior literature, we found that the association between new COVID-19 case counts and collective attention—one of our ancillary findings—was consistent with prior studies, including the impact of new daily cases on Australian tweets [11] as well as the impact of daily COVID-19 incidence on Reddit posts and comments across the United Kingdom, the United States, Canada, and Italy [12].

Limitations and Strengths

Twitter users may not be representative of the Canadian general population; therefore, our results may not be generalizable to the average Canadian. However, as of 2018, more than 15 million Canadians were classified as regular Twitter users (ie, use at least once per month) and represent a significant proportion of the 37 million members of the Canadian population [42]. One study found that North American Twitter users were younger, were more educated, and had higher income compared to the general population, but noted that their views were largely similar to the general population, except for their tendency to believe in the existence of gender and racial inequalities, which were lower in the general population [43]. In light of this information, we can interpret our findings as generalizable to a large portion of Canadians, especially for those who are younger, are more educated, have higher socioeconomic status, and tend to be more socially progressive.

Since our collection of tweets were based on keywords, there may be tweets that only contain less popular COVID-19–related keywords, such as “covidiots” or “antimask,” but do not contain common words, such as “COVID19” or “coronavirus.” While VADER has been specifically validated to analyze the sentiment in social media text, it is restricted to English-only tweets, and tweets written in other languages were not analyzed in our study. Finally, our study was not able to disentangle the separate effects of masking, social distancing, and travel restrictions, since (1) they all had similar start dates, (2) they were in effect for most of the study period, and (3) these restrictions were not lifted before the end of the study period. The overlapping nature of these restrictions limited our ability to investigate the unique contribution of their respective effects on public opinion.

Strengths of our study include the following:

1. The use of a multivariate statistical method to disentangle the effects of different pandemic restrictions; this provided stronger evidence for inference compared to prior literature studies that were largely descriptive in nature, which focused on documenting the tweet frequency and sentiment that coincided with COVID-19–related events [3,4,11].
2. Our study demonstrated the feasibility of using sentiment analysis to evaluate the impact of public health restrictions on public opinion, which can provide a relatively rapid and low-cost method to evaluate the impact of public health interventions compared to survey research.
3. We developed a novel approach of using the Gini index to measure sentiment polarization, where the index has been previously limited in its use as a measure of income disparity. Future studies may rely on the Gini index as a measure of sentiment polarization or level of disagreement.
4. Compared to prior studies that tended to focus only on the association between COVID-19–related events and collective attention, as measured by tweet frequency, our study examined the effect of restrictions on multiple dimensions of public opinion, including collective attention, aggregate sentiment, and level of disagreement, which provides a more holistic perspective of public opinion compared to single-measure studies.

Conclusions

Our study demonstrates the feasibility of combining sentiment analysis of social media text with dynamic regression models to understand the relationship between the introduction of COVID-19 restrictions and changes in public opinion over time, which provides a rapid and flexible method of evaluating the public response to large-scale restrictions. Our study also offers useful insights on the public opinion of COVID-19 restrictions; specifically, we showed that the impact of restriction on public opinion was contextually driven (eg, business closures were better tolerated with higher COVID-19 case counts), and while school closures and other restrictions generated increased collective attention, they did not have an effect on aggregate sentiment or the level of disagreement. Partial lockdowns were associated with better public response (ie, higher number of tweets with net positive sentiment and lower levels of disagreement) compared to province-wide lockdowns. This information can help public health practitioners anticipate public response to future pandemic restrictions and ensure adequate resources are dedicated to addressing increases in negative sentiment and levels of disagreement in the face of scientifically informed, but controversial, restrictions.

Acknowledgments

The study was funded by the Toronto COVID-19 Action Initiative grant (Fund No. 508849) from the University of Toronto, Ontario, Canada.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Timeline of key COVID-19–related events in Ontario from March to October 2020.
[DOCX File , 23 KB - jmir_v23i8e28716_app1.docx]

References

1. Rufai SR, Bunce C. World leaders' usage of Twitter in response to the COVID-19 pandemic: a content analysis. J Public Health (Oxf) 2020 Aug 18;42(3):510-516 [FREE Full text] [doi: 10.1093/pubmed/fdaa049] [Medline: 32309854]
2. Tsao S, Chen H, Tisseverasinghe T, Yang Y, Li L, Butt ZA. What social media told us in the time of COVID-19: A scoping review. Lancet Digit Health 2021 Mar;3(3):e175-e194 [FREE Full text] [doi: 10.1016/S2589-7500(20)30315-0] [Medline: 33518503]
3. Valdez D, Ten Thij M, Bathina K, Rutter LA, Bollen J. Social media insights into US mental health during the COVID-19 pandemic: Longitudinal analysis of Twitter data. J Med Internet Res 2020 Dec 14;22(12):e21418 [FREE Full text] [doi: 10.2196/21418] [Medline: 33284783]
4. Chandrasekaran R, Mehta V, Valkunde T, Moustakas E. Topics, trends, and sentiments of tweets about the COVID-19 pandemic: Temporal infoveillance study. J Med Internet Res 2020 Oct 23;22(10):e22624 [FREE Full text] [doi: 10.2196/22624] [Medline: 33006937]
5. Hou K, Hou T, Cai L. Public attention about COVID-19 on social media: An investigation based on data mining and text analysis. Pers Individ Dif 2021 Jun;175:110701 [FREE Full text] [doi: 10.1016/j.paid.2021.110701] [Medline: 33536695]
6. Zhu B, Zheng X, Liu H, Li J, Wang P. Analysis of spatiotemporal characteristics of big data on social media sentiment with COVID-19 epidemic topics. Chaos Solitons Fractals 2020 Nov;140:110123 [FREE Full text] [doi: 10.1016/j.chaos.2020.110123] [Medline: 32834635]
7. Nguyen TT, Criss S, Dwivedi P, Huang D, Keralis J, Hsu E, et al. Exploring US shifts in anti-Asian sentiment with the emergence of COVID-19. Int J Environ Res Public Health 2020 Sep 25;17(19):7032 [FREE Full text] [doi: 10.3390/ijerph17197032] [Medline: 32993005]
8. Lwin MO, Lu J, Sheldenkar A, Schulz PJ, Shin W, Gupta R, et al. Global sentiments surrounding the COVID-19 pandemic on Twitter: Analysis of Twitter trends. JMIR Public Health Surveill 2020 May 22;6(2):e19447 [FREE Full text] [doi: 10.2196/19447] [Medline: 32412418]
9. Kurten S, Beullens K. #Coronavirus: Monitoring the Belgian Twitter discourse on the severe acute respiratory syndrome coronavirus 2 pandemic. Cyberpsychol Behav Soc Netw 2021 Feb;24(2):117-122. [doi: 10.1089/cyber.2020.0341] [Medline: 32857607]
10. Wang H, Li Y, Hutch M, Naidech A, Luo Y. Using Tweets to understand how COVID-19-related health beliefs are affected in the age of social media: Twitter data analysis study. J Med Internet Res 2021 Feb 22;23(2):e26302 [FREE Full text] [doi: 10.2196/26302] [Medline: 33529155]
11. Yigitcanlar T, Kankanamge N, Preston A, Gill PS, Rezayee M, Ostadnia M, et al. How can social media analytics assist authorities in pandemic-related policy decisions? Insights from Australian states and territories. Health Inf Sci Syst 2020 Dec;8(1):37 [FREE Full text] [doi: 10.1007/s13755-020-00121-9] [Medline: 33078073]
12. Gozzi N, Tizzani M, Starnini M, Ciulla F, Paolotti D, Panisson A, et al. Collective response to media coverage of the COVID-19 pandemic on Reddit and Wikipedia: Mixed-methods analysis. J Med Internet Res 2020 Oct 12;22(10):e21597 [FREE Full text] [doi: 10.2196/21597] [Medline: 32960775]
13. Lewandowsky S, Jetter M, Ecker UKH. Using the president's tweets to understand political diversion in the age of social media. Nat Commun 2020 Nov 10;11(1):5764 [FREE Full text] [doi: 10.1038/s41467-020-19644-6] [Medline: 33173060]
14. Banda J, Tekumalla R, Wang G, Yu J, Liu T, Ding Y. A large-scale COVID-19 Twitter chatter dataset for open scientific research -- An international collaboration. ArXiv. Preprint posted online on November 13, 2020. [FREE Full text] [doi: 10.3390/epidemiologia2030024]
15. Tekumalla R, Banda JM. Social Media Mining Toolkit (SMMT). Genomics Inform 2020 Jun 15;18(2):e16 [FREE Full text] [doi: 10.5808/GI.2020.18.2.e16] [Medline: 32634870]
16. GeoNames. URL: https://www.geonames.org/ [accessed 2021-07-23]
17. Chumlab. COVID-19 Project. GitHub. 2020. URL: https://github.com/chumlab/covid19project [accessed 2021-03-03]
18. Hutto C, Gilbert E. VADER: A parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the 8th International AAAI Conference on Weblogs and Social Media. 2014 Presented at: 8th International AAAI Conference on Weblogs and Social Media; June 1-4, 2014; Ann Arbor, MI p. 216-225 URL: https://ojs.aaai.org/index.php/ICWSM/article/view/14550/14399
19. Botchway R, Jibril A, Kwarteng M, Chovancova M, Oplatková Z. A review of social media posts from UniCredit bank in Europe: A sentiment analysis approach. In: Proceedings of the 3rd International Conference on Business and Information Management. 2019 Sep 12 Presented at: 3rd International Conference on Business and Information Management; September 12-14, 2019; Paris, France p. 74-79. [doi: 10.1145/3361785.3361814]
20. Crocamo C, Viviani M, Famiglini L, Bartoli F, Pasi G, Carrà G. Surveilling COVID-19 emotional contagion on Twitter by sentiment analysis. Eur Psychiatry 2021 Feb 03;64(1):e17 [FREE Full text] [doi: 10.1192/j.eurpsy.2021.3] [Medline: 33531097]
21. Elbagir S, Yang J. Twitter sentiment analysis using Natural Language Toolkit and VADER sentiment. In: Proceedings of the 27th International MultiConference of Engineers and Computer Scientists. 2019 Presented at: 27th International MultiConference of Engineers and Computer Scientists; March 13-15, 2019; Hong Kong, China p. 12-16 URL: http://www.iaeng.org/publication/IMECS2019/IMECS2019_pp12-16.pdf [doi: 10.1142/9789811215094_0005]

22. Sprenger TO, Tumasjan A, Sandner PG, Welpe IM. Tweets and trades: The information content of stock microblogs. Eur Financ Manage 2013 May 29;20(5):926-957. [doi: 10.1111/j.1468-036x.2013.12007.x]
23. Mao H, Counts S, Bollen J. Quantifying the Effects of Online Bullishness on International Financial Markets. European Central Bank. 2015 Jul. URL: https://www.ecb.europa.eu//pub/pdf/scpsps/ecbsp9.en.pdf [accessed 2021-03-03]
24. Simonoff J, Sparrow I. Predicting movie grosses: Winners and losers, blockbusters and sleepers. Chance 2000;13(3):15-24. [doi: 10.1080/09332480.2000.10542216]
25. Asur S, Huberman B. Predicting the future with social media. In: Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology. 2010 Presented at: IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology; August 31-September 3, 2010; Toronto, ON p. 492. [doi: 10.1109/wi-iat.2010.63]
26. Kovacevic M. Measurement of Inequality in Human Development – A Review. United Nations Development Programme. 2010 Nov. URL: http://hdr.undp.org/sites/default/files/hdrp_2010_35.pdf [accessed 2021-08-13]
27. Thomas V, Wang Y, Fan X. Measuring Education Inequality: Gini Coefficients of Education. Washington, DC: The World Bank; Nov 1999.
28. Hörcher D, Graham DJ. The Gini index of demand imbalances in public transport. Transportation 2020 Sep 24:1-24 [FREE Full text] [doi: 10.1007/s11116-020-10138-4]
29. Manek AS, Shenoy PD, Mohan MC, Venugopal KR. Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier. World Wide Web 2017;20:135-154. [doi: 10.1007/s11280-015-0381-x]
30. COVID-19 intervention timeline in Canada. Canadian Institute for Health Information. 2021. URL: https://www.cihi.ca/en/covid-19-intervention-timeline-in-canada [accessed 2021-03-03]
31. Enhanced Epidemiological Summary: COVID-19 in Ontario: A Summary of Wave 1 Transmission Patterns and Case Identification. Public Health Ontario. 2020. URL: https://www.publichealthontario.ca/-/media/documents/ncov/epi/2020/08/covid-19-wave-1-transmission-patterns-epi-summary.pdf?la=en [accessed 2021-08-13]
32. Nielsen K. A timeline of COVID-19 in Ontario. Global News. 2020 Apr 24. URL: https://globalnews.ca/news/6859636/ontario-coronavirus-timeline/ [accessed 2021-03-03]
33. A timeline of events in Canada's fight against COVID-19. CP24. 2020 Dec 15. URL: https://www.cp24.com/news/a-timeline-of-events-in-canada-s-fight-against-covid-19-1.5231865 [accessed 2021-03-03]
34. The Daily: Canada's population estimates: Subprovincial areas, July 1, 2019. Statistics Canada. 2020 Feb 13. URL: https://www150.statcan.gc.ca/n1/daily-quotidien/200213/dq200213a-eng.htm [accessed 2021-04-05]
35. Slaughter G. Canadians prefer strict lockdowns over partial shutdowns for hard-hit regions: Nanos survey. CTV News. 2020 Dec 15. URL: https://www.ctvnews.ca/health/coronavirus/canadians-prefer-strict-lockdowns-over-partial-shutdowns-for-hard-hit-regions-nanos-survey-1.5231626 [accessed 2021-03-03]
36. Hu W. Real-time Twitter sentiment toward Thanksgiving and Christmas holidays. Soc Netw 2013 Apr;2:77-86. [doi: 10.4236/sn.2013.22009]
37. Jackson H. Canada could see "grotesque" spike in coronavirus cases after holidays: Expert. Global News. 2020 Dec 24. URL: https://globalnews.ca/news/7540186/canada-coronavirus-christmas-spike/ [accessed 2021-03-03]
38. CSSEGISandData. COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. GitHub. 2021. URL: https://github.com/CSSEGISandData/COVID-19 [accessed 2021-03-03]
39. Jebb AT, Tay L, Wang W, Huang Q. Time series analysis for psychological research: Examining and forecasting change. Front Psychol 2015;6:727 [FREE Full text] [doi: 10.3389/fpsyg.2015.00727] [Medline: 26106341]
40. Pankratz A. Forecasting with Dynamic Regression Models. Hoboken, NJ: John Wiley & Sons; 2012.
41. Woodward W, Gray H, Elliott A. Applied Time Series Analysis With R. 2nd edition. Boca Raton, FL: CRC Press; 2017.
42. Slater M. By the numbers: Twitter Canada at Dx3 2018. Twitter. 2018 Mar 12. URL: https://blog.twitter.com/en_ca/topics/insights/2018/TwitterCanada_at_Dx3.html [accessed 2021-03-03]
43. Wojcik S, Hughes A. Sizing Up Twitter Users. Pew Research Center. 2019 Apr 24. URL: https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2019/04/twitter_opinions_4_18_final_clean.pdf [accessed 2021-03-03]

Abbreviations

AIC: Akaike information criterion
AR: autoregressive
ARIMA: autoregressive integrated moving average
MA: moving average
VADER: Valence Aware Dictionary and Sentiment Reasoner

Edited by G Eysenbach; submitted 11.03.21; peer-reviewed by H Wang, M Tizzani; comments to author 01.04.21; revised version received 09.04.21; accepted 26.04.21; published 25.08.21.

Please cite as:
Chum A, Nielsen A, Bellows Z, Farrell E, Durette PN, Banda JM, Cupchik G
Changes in Public Response Associated With Various COVID-19 Restrictions in Ontario, Canada: Observational Infoveillance Study Using Social Media Time Series Data
J Med Internet Res 2021;23(8):e28716
URL: https://www.jmir.org/2021/8/e28716
doi:10.2196/28716
PMID:34227996

©Antony Chum, Andrew Nielsen, Zachary Bellows, Eddie Farrell, Pierre-Nicolas Durette, Juan M Banda, Gerald Cupchik. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 25.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Implementation of Digital Monitoring Services During the COVID-19 Pandemic for Patients With Chronic Diseases: Design Science Approach

Luís Velez Lapão[1,2,3*], MSc, PhD; Mariana Peyroteo[2,4], MSc; Melanie Maia[1,5], MSc; Jorge Seixas[1], PhD; João Gregório[6], MSc, PhD; Miguel Mira da Silva[7], MSc, MBA, PhD; Bruno Heleno[2*], MD, PhD; Jorge César Correia[8*], MD, MSc

[1]Global Health and Tropical Medicine, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Lisbon, Portugal

[2]Comprehensive Health Research Center, NOVA Medical School, Universidade Nova de Lisboa, Lisbon, Portugal

[3]Research and Development Unit in Mechanical and Industrial Engineering (UNIDEMI), NOVA School of Science and Technology, Universidade Nova de Lisboa, Caparica, Portugal

[4]NOVA School of Science and Technology, Universidade Nova de Lisboa, Caparica, Portugal

[5]NOVA School of Social Sciences and Humanities, Universidade Nova de Lisboa, Lisbon, Portugal

[6]Research Center for Biosciences and Health Technologies, Universidade Lusófona de Humanidades e Tecnologias, Lisbon, Portugal

[7]Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal

[8]Unit of Patient Education, Division of Endocrinology, Diabetology, Nutrition and Patient Education, Department of Medicine, Geneva University Hospitals and University of Geneva, Geneva, Switzerland

*these authors contributed equally

Corresponding Author:
Luís Velez Lapão, MSc, PhD
Global Health and Tropical Medicine
Instituto de Higiene e Medicina Tropical
Universidade Nova de Lisboa
Rua da Junqueira, 100
Lisbon, 1349-008
Portugal
Phone: 351 213652600
Email: luis.lapao@ihmt.unl.pt

Abstract

Background: The COVID-19 pandemic is straining health systems and disrupting the delivery of health care services, in particular, for older adults and people with chronic conditions, who are particularly vulnerable to COVID-19 infection.

Objective: The aim of this project was to support primary health care provision with a digital health platform that will allow primary care physicians and nurses to remotely manage the care of patients with chronic diseases or COVID-19 infections.

Methods: For the rapid design and implementation of a digital platform to support primary health care services, we followed the Design Science implementation framework: (1) problem identification and motivation, (2) definition of the objectives aligned with goal-oriented care, (3) artefact design and development based on Scrum, (4) solution demonstration, (5) evaluation, and (6) communication.

Results: The digital platform was developed for the specific objectives of the project and successfully piloted in 3 primary health care centers in the Lisbon Health Region. Health professionals (n=53) were able to remotely manage their first patients safely and thoroughly, with high degrees of satisfaction.

Conclusions: Although still in the first steps of implementation, its positive uptake, by both health care providers and patients, is a promising result. There were several limitations including the low number of participating health care units. Further research is planned to deploy the platform to many more primary health care centers and evaluate the impact on patient’s health related outcomes.

(J Med Internet Res 2021;23(8):e24181) doi:10.2196/24181

KEYWORDS

primary healthcare; information systems; telemedicine; implementation; design science research; COVID-19; monitoring; chronic disease; elderly; digital health

Introduction

Since the first case of COVID-19 was diagnosed in Wuhan, China in October 2019, the virus quickly spread around the world to become a global pandemic and a public health emergency, as declared by the World Health Organization in March 2020 [1]. According to the latest available data, more than 180 countries have been affected, with over 178 million confirmed cases including more than 3 million deaths due to the virus [2].

Portugal has also been severely affected by this pandemic, with 420,629 confirmed COVID-19 infections and 6972 deaths by the end of 2020 [3,4]; however, thanks to several important public health measures (eg, closing schools, obligatory used of masks, broad application of testing, and closing of nonessential services) that were quickly put into place by health authorities, the situation was contained, and Portugal has been identified as a moderate success case in managing the pandemic [3]. The Portuguese health system benefits from its universal coverage (eg, all populations have free access to public health, primary, and hospital care in the national health service) and several public health units, which are strategically spread around the country, allowing easy access to care and facilitating contact tracing. Nevertheless, the transmission of the virus persisted and there was a significant increase in mortality rates, particularly among the population over 65 years of age [4] before the start of the vaccination campaign.

It has been observed that older adults and those with underlying health conditions are at increased risk of contracting COVID-19 and have an increasingly rapid and severe progression, often leading to death [5]. Moreover, according to a recent modeling study [6], approximately 1 in 5 individuals worldwide fall into this increased risk category. The increased mortality rate in these populations is expected to be reduced as the vaccination is rolled out, which began in December 2020 [7].

Another cause for concern associated with this pandemic is the severe and widespread disruption of prevention and treatment services for populations affected by chronic conditions. This was reflected in a survey conducted by the World Health Organization [8], involving 155 countries, in which shortages of services for treatment of hypertension (53%), cardiovascular emergencies (31%), for treatment of diabetes and diabetes-related complications (49%), and for cancer treatment (42%) were reported, due to COVID-19 primary health care services disruption. In Portugal, it is estimated that more than 200,000 surgeries and 2 million consultations were delayed due to the COVID-19 pandemic [3].

Several studies [9,10] have also identified the issue wherein people have delayed or avoided seeking medical care for life-threatening conditions, due to concerns about being exposed to COVID-19 in the hospital. Other indirect effects related to the pandemic that have negatively impacted population health were identified, such as an increase in domestic violence and psychological distress related to social isolation [11].

All these issues have led to a renewed focus on telehealth, as a means of providing care to both patients with COVID-19 infections and those requiring other routine clinical services without increasing the risk of potential exposure for patients, clinicians, and staff [12,13].

In this context, we sought to design a web-based digital platform to support primary health care services during the COVID-19 pandemic, by facilitating web-based consultations between primary health care teams and their patients, to guarantee appropriate care, promote adherence to treatment, provide counseling and psychological support.

We aim to share our efforts and experiences in creating a digital health service during a public health emergency (ie, in the midst of a pandemic). This project resulted from a collaboration between academia and health professionals and allowed for the timely deployment (within 3 months) of an innovative digital platform to improved access to health care during lockdown.

Methods

Research Method

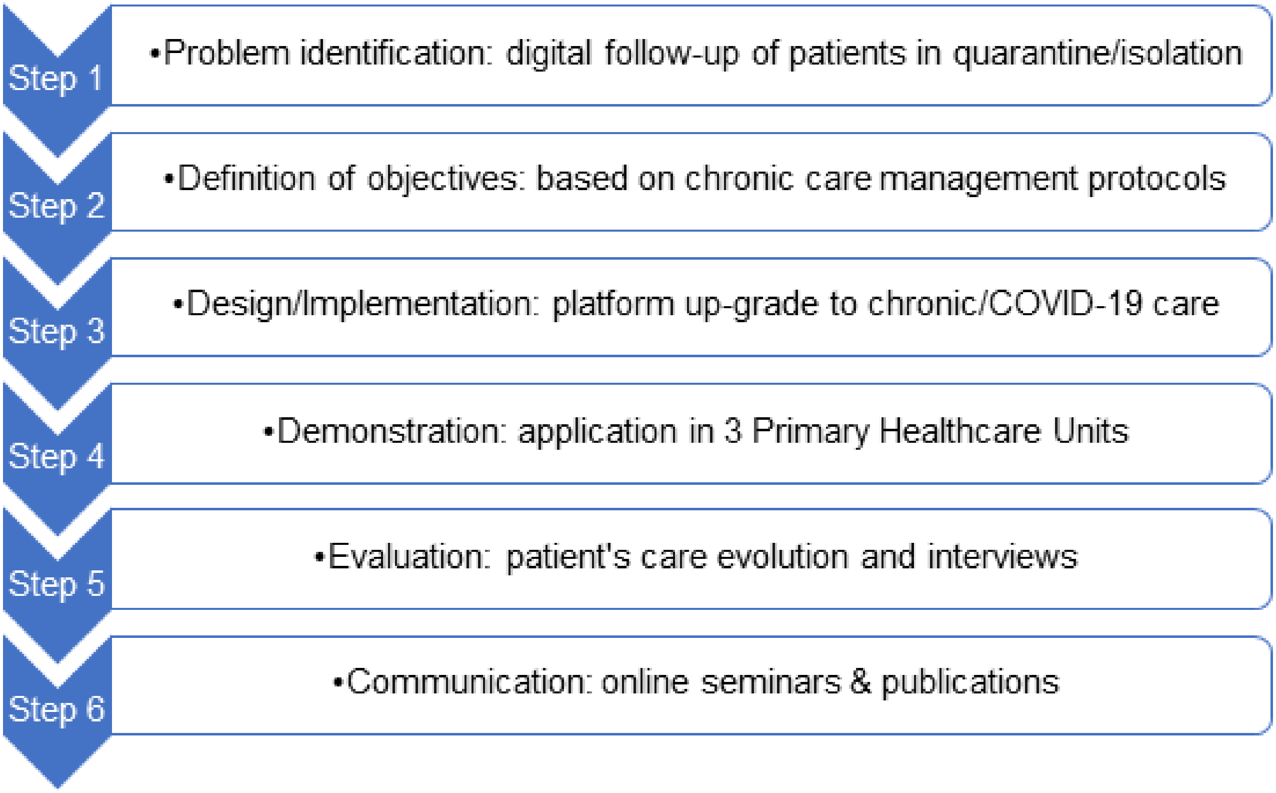
Overview

We designed, created, and implemented a digital platform with specific features in the shortest possible time under difficult circumstances imposed by the pandemic situation using the pace established by Scrum and in a close collaboration with the Primary Health Care professionals who would be using the platform.

We followed the Design Science Research Methodology [14,15], which is based on a process with 6 sequential steps (Figure 1). Design Science Research Methodology benefits from supporting the design of an artefact (the primary health care digital platform) to solve the identified problem of lacking the access to chronic patients.

Each step—problem definition, defining the objectives, and design and development, demonstration, evaluation and communication—was performed as a research task.

Figure 1. Design science research process.


Step 1
•Problem identification: digital follow-up of patients in quarantine/isolation
Step 2
•Definition of objectives: based on chronic care management protocols
Step 3
•Design/Implementation: platform up-grade to chronic/COVID-19 care
Step 4
•Demonstration: application in 3 Primary Healthcare Units
Step 5
•Evaluation: patient's care evolution and interviews
Step 6
•Communication: online seminars & publications

Problem Definition

A set of interviews with general practitioners, nurses, and clinical secretaries helped identify both the critical chronic care processes (eg, consultations frequency and service levels depending on health risks) and the major communication problems between chronic patients and health care professionals (eg, the role of each health professional, how to clearly communicate therapies, and how to improve adhesion), and moreover, in the pandemic, how to access to chronic patients at home.

Defining the Objectives

The aim of this activity was to identify the digital care services required by the patients when they interact with their primary health care providers, specifically during lockdown due to the pandemic. This helped to inform the functionalities that the health professionals felt were lacking in care, supported by functionalities that were intended to mimic the work processes of the goal-oriented care methodology for primary health care provision [16]. Goal-oriented care is a patient communication strategy that aims to increase patient responsibility by engaging and involving them more in their therapeutic decision-making process.

Design and Development

The design and development of the digital platform was based on the goal-oriented care methodology for primary health care provision [16] and on an existing instantiation that provides digital pharmaceutical services by interacting with chronic patients at home [15]. The use of goal-oriented care principles (ie, creating mechanisms for sharing decision making with patients) allowed us to design service characteristics and functionalities specifically required by chronic patients (eg, the possibility of teleconsultations that enable more frequent consultations as demanded by the patients' needs and choice of monitoring indicators for more comprehensive evaluation of the patient)). This task was based on Scrum [17] using 1-week sprints with Trello [18]. Scrum is an agile project management method that collects feedback from end-users at the end of each sprint, and new features are prioritized often to reduce risk and extract maximum value [17]. Furthermore, in this project, after the end of each sprint, a new working version of the platform was published and immediately tested by the general practitioners.

Demonstration

The platform was then applied in practice engaging the 3 family health units' professionals for 4 weeks. This allowed us to test online medical and nursing teleconsultations as well as disease monitoring through patient-shared health data.

Evaluation

Data from the demonstration were collected. The platform utility was then evaluated based on the teleconsultations effectiveness and patient's adherence to care, as well as major bottlenecks in users' experiences. The evaluation also provided more information about the usefulness of the digital service and identified missing features for implementation in the next design cycle. This step will be fully completed after 12 months of the deployment of the platform.

Communication

This activity aims to disseminate the project's results. This activity will be performed throughout the duration of the research project. It includes webinars, oral communications, papers published in conference proceedings within the fields of study and papers to be published in peer reviewed journals.

The project received Ethics approval from the Lisbon Health Region Ethics Commission, and in accordance with international security (ISO 27001) and privacy (General Data Protection Regulation) standards.

Health Care Setting

In Portugal, primary health care is largely provided by a publicly funded National Health Service System. Regional Health Authorities ensure access through primary care through primary care trusts. These trusts coordinate different types of primary care practices (family health units, public health services, home care services, and allied health professional services) and support units (the executive director, a clinical council, and a management unit) to around 150,000 to 400,000 citizens. Primary care is provided by smaller practices (family health units), generally composed of teams of 15 to 30 health care providers, with 6 to 11 family physicians, an equal number of family nurses, and a smaller number of administrative staff and trainees. These practices serve a fixed list of around 15,000 citizens [12].

There is a mandatory use of electronic health records and electronic prescription; however, the use of videoconferencing is not widespread, mostly because health care units do not have the proper technical equipment for teleconsultations (eg, no cameras are available).

The primary health care digital platform was piloted during the month of July 2020, in 3 family health units (USF for *Unidade de Saúde Familiar* in Portuguese) in the Lisbon Health Region. A USF is an operating unit of primary health care centers with functional and technical autonomy, where patients are closely managed and monitored by a group of primary health care professionals, thus representing an ideal site for our pilot project.

The pilot project included 3 USFs that were purposely selected. *USF Jardim dos Plátanos* is a model A family health unit (those under a commissioning agreement for services delivered), located in a suburban area of Greater Lisbon. At the end of 2019, there were 15,011 registered users and the team consisted of 20 professionals: 8 doctors, 7 nurses, and 5 clinical secretaries. The USF also included family medicine residents and assistant nurses and specialists, as well as technical assistants. *USF das Conchas* is also a model A family health unit, located in Lisbon. There were 13,750 registered patients, 8 doctors, 7 nurses, and 5 clinical secretaries. *USF Ribeirinha* is a model A family health unit, located south of Lisbon, on the south bank of river Tagus. There were 14,962 registered patients, 9 physicians (and 10 residents), 8 nurses, and 6 clinical secretaries.

These health care units were selected for their previous participation in innovative projects addressing the organization of primary health care digital services and their strategic location in a region with a high population density and with a high number of COVID-19 infections and because of the high prevalence of chronic conditions in registered patients.

Participants

The new digital service was co-designed (eg, functionalities and design options) involving patients and health care providers.

Patients

This research included populations that were vulnerable to the COVID-19 infection, namely older adults and those affected by chronic conditions as recommended by health authorities [19-21]. Participants were recruited by physicians if they were older than 60 years of age or had chronic health conditions such as chronic obstructive pulmonary disease, asthma, chronic heart failure, chronic ischemic heart disease, chronic cerebrovascular disease, diabetes, cancer, or rheumatic conditions under disease modifying drugs. For this initial pilot study, patients with low technological literacy or who did not give their consent were excluded. Patients were contacted by the physicians and provided an email, which allowed them to receive the link and the password to access the digital portal. They were also given an email address to contact the research team if necessary.

Health Care Providers

We included physicians and nurses in these 3 practices who were willing to participate in the study. All the teams were led by their USF coordinator. Regular team work meetings (eg, each fortnight or each month) were organized with providers to explain and test the digital platform and, in which, we received their feedback and suggestions.

Digital Health Service Platform

The digital platform was developed based on an existing system from previous eHealth research projects conducted by members of the research team, namely ePharmaCare [15] and HAITool [22]. The original ePharmaCare platform included the following main functionalities: a record of patients' diseases and medicines, the user messaging inbox and the possibility of chatting between patients and health professionals, records of patients' physiological and biochemical data (with graphics), a tool for monitoring mental health and quality of life, and the possibility of pharmacists issuing a monthly report to send to the general practitioner. For each type of user profile—professional or patient—there was a home dashboard set up with the main features for its use, with direct links to the information they needed most.

Design Science Research Methodology was used to adapt the platform to the specific objective of the project—namely to allow primary health care physicians and nurses to monitor patients with chronic diseases and those with COVID-19 infections under quarantine [15-23]—by the research team under a participatory process that involved researchers, digital health experts, primary health care professionals, and patients. The health care team of each participating site worked in close collaboration with the researchers (monthly meetings were scheduled for analyzing the advances and collecting suggestions to improve the platform), in order to ensure that the platform responded to their needs (thus guaranteeing adoption and sustainability).

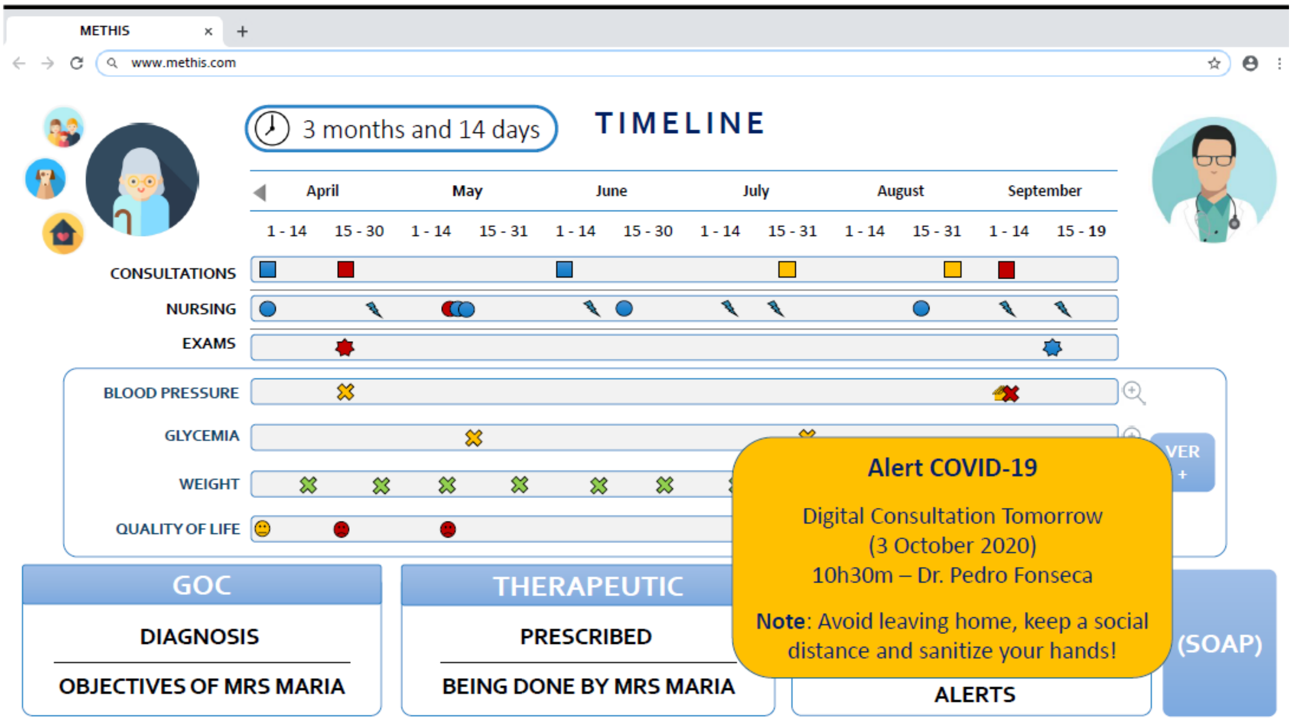
In terms of architecture, the primary health care digital platform (METHIS: Multimorbidity Management Health Information

System) is a web-based app with 3 relational databases using PostgreSQL. There is a staging database to enable proper testing before production, a production database to collect USF data, and a research database, with anonymized data from the production database. The platform was also integrated with Zoom (Zoom Inc) to allow for teleconsultations.

We expected that the digital platform would work as a chronic disease management portal, where health care professionals, patients, and caretakers could manage patients' chronic conditions. It was built to allow both synchronous (consultation videoconferences) and asynchronous communication (patients can record health goals, vital signs, condition-specific scales or general health-related quality of life instruments, and results of investigations).

The digital platform promotes and facilitates communication, by including integrated views of the patient's demographic information, chronic conditions, medication, exams, and personal goals—inspired by goal-oriented care [16] and displayed with innovative layouts and graphics (Figure 2).

Figure 2. Web-based user interface of the platform (example from a fictional case).



Alert COVID-19

Digital Consultation Tomorrow
(3 October 2020)
10h30m – Dr. Pedro Fonseca

Note: Avoid leaving home, keep a social distance and sanitize your hands!

The platform presents a set of functionalities that allow scheduling and management of consultations using telephone (voice call) or via digital videoconferencing using Zoom. Teleconsultation scheduling requires only 3 clicks (to select the date, hour, and confirm the consultation that sends the alert to the patient). The platform is also a decision-support system, displaying alerts if patient inputs are out of the specified range upon automonitoring (eg, blood pressure measurement or blood glucose levels) and for drug interactions. The platform also manages and informs patients regarding COVID-19 infection and safety measures. All patient-related health information is processed in a secure manner in accordance with General Data Protection Regulation rules.

Research Team

The project was proposed and implemented by a multidisciplinary team of 9 researchers, mainly from *Universidade Nova de Lisboa* (with the coordination), including also researchers from 3 other institutions, namely *Instituto Superior Técnico* and *Universidade Lusófona* in Portugal and the Geneva University Hospitals in Switzerland.

The research team comprised 3 physicians (professor of general practice and general practitioner at one of the participating practices, professor of global health and tropical medicine, and senior researcher in chronic disease management and therapeutic patient education), 1 nurse and PhD student, 1 professor of community pharmacy, 1 expert in health communication and PhD student, 1 professor of information systems, 1 expert in digital platforms and PhD student, and 1 professor of health information systems as the team leader (and leader of ePharmaCare [15]).

Part of the team was focused on the development of the digital platform and on direct contact with the family health units (tests and discussions), while the other part of the team provided problem-solving expertise (more often during Scrum meetings but also when specific problems required additional contact).

Results

Overview

The project started in May 2020 and by the beginning of July 2020 the platform had been fully deployed (within 9 weeks) and delivered to the 3 family health unit coordinators. We provided 2 tablets (with cameras for teleconsultations) to each USF. A kick-off session was organized with each USF to facilitate the adoption of the platform and address any questions.

Every USF performed their first consultation with a real patient during this meeting (the patients were contacted beforehand).

The research problem had been previously defined and the project had identified that there were barriers to communication with chronic patients, such as patients' lack of mobility and lack of information between the consultations [15].

To define and validate the objectives of a solution (eg, what were the functionalities to be implemented in the digital portal to help link health professionals with patients), focus groups in each of the 3 USFs were organized, and a set of functionalities were specified for integration into the digital platform. All the functionalities that were proposed were validated in the Scrum meetings.

The Scrum meetings took place every week (Monday morning, 60-minute maximum duration) and were supported by Trello to manage the information about the sprints. The meetings started with analyzing the previous week's sprint, validating the activities that were accomplished as planned, discussing feedback from end-users, and defining a new delivery date for the activities that were, for any reason, not finished as expected. In the second part of the meeting, a new sprint that took into the project timetable into consideration, including who was responsible for each activity, was defined.

Next, the process of designing the new digital platform was developed in collaboration with the physicians, nurses, and clinical secretaries of the 3 USFs. Often the physicians of our team would help clarify requirements for a functionality, and the digital platform team would analyze how it could be implemented. In every Scrum meeting, validated functionalities were assigned to a timeframe, to be developed and integrated into the platform. Nonvalidated functionalities were transferred to staging, for future consideration. After the first set of meetings, it was decided that the platform would include 5 priority features.

Features

Teleconsultations

This feature allows appointment to be easily booked for teleconsultations. Prior to the appointment, a message is sent to both physicians and patients with the schedule and the unique link to the teleconsultation. It also included a SOAP (Subjective, Objective, Assessment, Plan) free-text structure, as demanded in primary health care settings [24]. If part of this information was of interest to the patient, it could be also shared with them, with a simple click to validate this option.

Monitoring Patient Data

This feature allows monitoring of patients' biochemical and physiologic data as well as prescribed medicines, with several graphic displays available to facilitate interpretation and follow-up. This data can be uploaded by the patient (at home or after examination at the community pharmacy) or by the physician (at the consultation).

Patient Participation

This feature allows patients to upload daily or regular physiologic and biochemical data that can be later visualized by physicians or the patients themselves. Patients can also ask questions and share information with the health professionals via a chat system, in addition to receiving information sent by the health professionals.

Alerts

The platform has new smart components, including algorithms that allow a set of alerts to improve physicians' and nurses' responses to a potential negative evolution in patients' data. As an example, all physiological and biochemical indicators have healthy limits included in the system, and whenever the indicator is outside these limits, or whenever this indicator is n-times outside the predetermined normal interval defined by the physician, it issues an alert.

Medicine Management

Using the Portuguese medicine database webservices provided by the National Authority for Medicines and Health Products (INFARMED), we included a module for therapeutic management. This module allows professionals or patients, to input the name of their medication (generic or commercial name) and dosage to yield a therapeutic profile. A feature of this profile is to automatically calculate the end date for each medication package prescribed, assisting with medicine management and therapeutic adherence issues. This feature, alongside the participation of a specialist pharmacist, will allow for different types of medication reviews and enhanced pharmaceutical care. This last function was inherited from the ePharmaCare project.

Demonstration and Evaluation

The demonstration took place after a meeting with USF teams in July 2020. There was a positive uptake by health care providers and patients in a truly short term. As of December 31, 2020, there were 53 health care providers (37 physicians and 16 nurses) from 3 USFs engaged and using this platform to follow 35 patients (Table 1).

In order to provide the first evaluation of the digital care provided, during the sequential meetings and first tests, we observed, listened, and recorded health care providers talking about their experiences and difficulties while using the platform.

At 3 kick-off meetings with USFs (in the beginning of July), the physicians and nurses showed marked enthusiasm. In all occasions, they immediately took the new tablets with the system and started using them without almost any help from the research team.

Interestingly, in the first presentation of the digital platform by the researchers, during which physicians were told that they could define functionalities that they needed to be added to the platform, 1 physician enquired whether they could really ask for additional functionalities for the system.

We observed that the first patient at USF *Jardim dos Plátanos*, a retired lawyer of 83 years of age (who provided written consent), acknowledged the reception of the email with the log-in and password. Soon after, he managed to initiate the teleconsultation easily, without any help from the physician.

A physician, during her first consultation and while talking with the patient, managed to effortlessly book another appointment (scheduled for after the holidays, 1 month afterwards) and requested that the patient enter his weight and arterial pressure every week in the platform. The patient said he would try.

Several physicians and nurses pointed-out the user-friendliness of the digital platform, and the reduced number of clicks necessary to perform the basic functions (eg, a teleconsultation with a patient requires only 3 clicks including log-in to the platform).

Table 1. Description of the number of health care provider and patients per unit until December 31.

USF[a]	Physicians, n	Nurses, n	Patients, n	Teleconsultations, n
A	15	7	11	12
B	13	7	7	8
C	9	2	17	21
Total	37	16	35	41

[a] USF: family health unit (*Unidade de SaúdeFamiliar*).

In 1 USF, the 2 physicians mentioned that the platform could also be used to follow pregnant women, since the protocols are well-known, and it would reduce their COVID-19 infection risk (and they already have used it for that purpose).

It was observed that each USF's team preferentially used different functionalities of the platform that were in line with the work and organizational processes of each site. Indeed, while 1 USF preferred to use the platform mostly for teleconsultation, another used it more for the recording and monitoring of patient data and another to focus on engaging more digitally literate patients. Understanding this behavior could be a possible focus for a future study.

Several physicians mentioned difficulty getting the tablet (often kept in a lockable drawer) when they were busy with consultations in order to perform teleconsultations. At USF *Ribeirinha*, a clinical secretary was called to help tackle this barrier.

The USF coordinator at *Conchas* was naturally skeptical at the beginning, but soon after using the platform, he accepted the challenge and assumed the leadership of using the system in his USF. In the meantime, we had a meeting at this USF with the National eHealth Institute, where this coordinator played an important role in promoting the value of the platform.

An important measure of some difficulties was the slow rate of adding new patients. This was, the health professionals said, in part, limited by the holiday season and related absences of many health care providers and patients. Logistic difficulties due to COVID-19 were also a barrier.

Important features of the digital platform that had not been possible before were the ability to contact and see the patients (not possible in the telephone calls) and to make appointments with reminders for upcoming teleconsultations. We were told that several patients had already started entering their own data onto the platform.

Regarding dissemination, in addition to an oral communication presented to a special COVID-19 session [25], a press release was also sent to the media that resulted in an interview in the Portuguese Public Television. More importantly, this paper fulfills the methodological aim.

Discussion

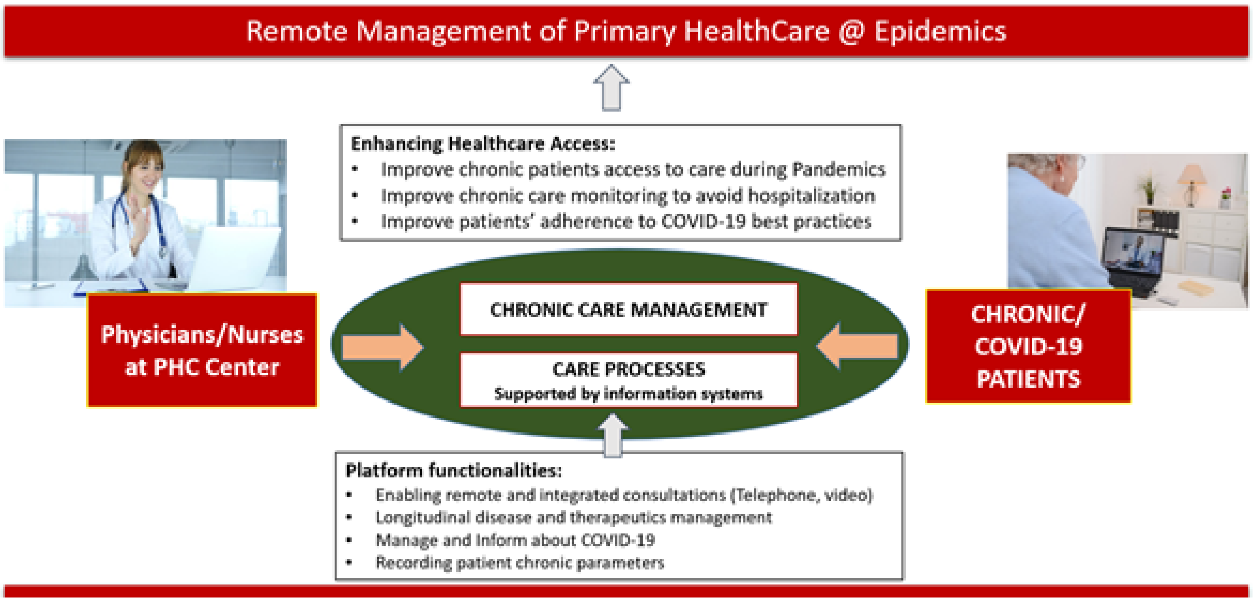
The main objectives of this rapid implementation project were attained: a digital platform was designed, implemented, and demonstrated in 3 different USFs, with several health care providers already using it to manage a set of their patients remotely (Figure 3).

The rapid implementation (in 9 weeks and within the 3-month limit, as required by the funding entity) was focused on the development of the platform, with main organizational implications, such as communication with patients and the number of interactions (many now through digital media).

Tablets were chosen for 4 reasons: camera, easy access to the internet, inexpensive, and easy to use. They can be used by different physicians and nurses throughout the day and provided the capabilities for performing both teleconsultations and monitoring. The results thus far have provided valuable proof-of-concept feedback and proved that health care professionals in the USFs could provide remote consultations to their patients during the COVID-19 pandemic using our digital health service.

We observed that health care providers were able to safely communicate with and prescribe appropriate treatment to their patients, as well as inform them about the COVID-19 related protective measures and answer any other queries they might have (eg, mental health). Yet, we did not have the opportunity to assess patients either on quarantine or in isolation as, at the time of implementation, the rate of COVID-19 infected patients was still relatively low in Portugal (approximately 0.8% [26]), and the physicians at the 3 USFs reported 1 or fewer infections in their populations. However, physicians were reportedly very busy taking care of chronic patients, using either telephone calls or emails, as most USF stopped attending patients in person. The competition, usual in any innovative process, between the 2 solutions (telephone vs digital platform) seemed to tend toward the simplest option of the telephone call, as the physicians claimed to be overloaded. Yet, 2 to 3 health professionals from each USF were enthusiastically looking for opportunities for using the digital platform, expecting to obtain the most relevant patient information from their patients with more digital proficiency.

Figure 3. Overview of the project. PHC: primary health care.


Remote Management of Primary HealthCare @ Epidemics
Enhancing Healthcare Access:
Improve chronic patients access to care during Pandemics
Improve chronic care monitoring to avoid hospitalization
Improve patients' adherence to COVID-19 best practices
Physicians/Nurses at PHC Center
CHRONIC CARE MANAGEMENT
CARE PROCESSES Supported by information systems
CHRONIC/ COVID-19 PATIENTS
Platform functionalities:
Enabling remote and integrated consultations (Telephone, video)
Longitudinal disease and therapeutics management
Manage and Inform about COVID-19
Recording patient chronic parameters

There are similar projects that are currently being developed in countries all over the world, such as in Canada [1,2], the United Kingdom [3], and China [4-6], with promising results that suggest that, if implemented and delivered appropriately, digital health solutions can effectively reduce the burden on hospitals, prevent overcrowding, reduce the risk of cross-infection, and relieve patient anxiety.

Although we have not completed the evaluation of our service, there is increasing evidence that clinical consultations conducted through a video link tend to be associated with high satisfaction among patients and staff, with no difference in disease progression and lower costs compared with traditional clinic-based care [10]. In fact, several recommendations are being developed to help primary health care professionals remotely provide COVID-19–related consultations [11]. One aspect to be taken into consideration is the actual support of the clinical secretary in managing the first level of communication with the patients as envisaged in the beginning of the project (ie, the digital portal includes the role of the clinical secretary), which is a functionality that most physicians did not take into consideration.

Nevertheless, there is still a lack of national and local guidelines regarding the remote management of patients using digital health solutions, and urgent research is required in this field. Our research efforts may contribute, albeit indirectly, to solving this problem. Other risks of using this type of platform include the possibility of misdiagnoses, equipment malfunction, and privacy breaches [8]. Other issues that need to be addressed include inadequate training of health care providers of skills in dealing with remote consultations to provide safe and effective patient care [9]. And, of course, the widespread use of resources, such as those demonstrated herein, is always dependent on patients' adequate technological literacy and internet access [27].

Other researchers have also highlighted the importance of implementing digital solutions without further fragmenting the existing landscape of care [13]. This requires concerted efforts between policy makers, researchers, and health professionals that take into account feedback from the patients. In Portugal, the primary health care activities are annually commissioned, to target the number of chronic care procedures to be accomplished but not the comprehensive well-being of the patient [12].

This digital platform will also be a valuable source for epidemiological data. We are also currently discussing different functionalities that can be used for research purposes (eg, to evaluate the effectiveness of a digital program for diabetic patient education using a virtual assistant, or the follow-up of disease-specific cohorts). The project team was recently contacted by a set of health care centers inquiring how to be included in the study, and by African Portuguese-speaking country partners, as they have shown interest in this platform for advancing universal coverage in their countries.

Although still in the first steps of project dissemination, we are seeing promising results with a positive acceptance by health care providers and patients. The primary health care units now have a digital platform for following and monitoring chronic patients, which can be use in the eventuality of subsequent waves of COVID-19 infections or during the flu season.

This study had several limitations. Most notably, the results may not be generalizable, as only 3 health units participated in the study, and it was developed under the restrictions of COVID-19 pandemic (eg, several meetings with health professionals were conducted via videoconference). Further research is required to evaluate the impact on patient health-related outcomes. We believe that this platform could be scaled-up to many more primary health care centers in Portugal.

Acknowledgments

This project (PrimaryCare@COVID-19: Digital Platform for Supporting Chronic Patients and COVID-19 consultations and Monitoring in Primary Care) was funded by the *Fundação para a Ciência e Tecnologia* (Ministry of Science of Technology, Portugal) under the Research4COVID-19 program.

We would like to acknowledge the contribution and engagement from the health professionals from the 3 health units: *Jardim dos Plátanos*, *Ribeirinha*, and *das Conchas*.

Conflicts of Interest

None declared.

References

1. Ghebreyesus TA. WHO director general's opening remarks at the media briefing on COVID-19. World Health Organization. Geneva; 2020 Mar 11. URL: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 [accessed 2021-08-08]
2. WHO coronavirus disease (COVID-19) dashboard. World Health Organization. Geneva: World Health Organization; 2021 Jun 23. URL: https://covid19.who.int [accessed 2021-08-08]
3. Oliveira M, Fernandes C. Managing the coronavirus pandemic in Portugal: a step-by-step adjustment of health and social services. Psychol Trauma 2020 Jul;12(5):536-538. [doi: 10.1037/tra0000879] [Medline: 32525369]
4. Nogueira PJ, Nobre MDA, Nicola PJ, Furtado C, Vaz Carneiro A. Excess mortality estimation during the COVID-19 pandemic: preliminary data from Portugal. Acta Med Port 2020 Jun 01;33(6):376-383. [doi: 10.20344/amp.13928] [Medline: 32343650]
5. Sanyaolu A, Okorie C, Marinkovic A, Patidar R, Younis K, Desai P, et al. Comorbidity and its impact on patients with COVID-19. SN Compr Clin Med 2020 Jun 25:1-8 [FREE Full text] [doi: 10.1007/s42399-020-00363-4] [Medline: 32838147]
6. Clark A, Jit M, Warren-Gash C, Guthrie B, Wang HHX, Mercer SW, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 working group. Global, regional, and national estimates of the population at increased risk of severe COVID-19 due to underlying health conditions in 2020: a modelling study. Lancet Glob Health 2020 Aug;8(8):e1003-e1017 [FREE Full text] [doi: 10.1016/S2214-109X(20)30264-3] [Medline: 32553130]
7. Viana J, van Dorp CH, Nunes A, Gomes M, van Boven M, Kretzschmar ME, et al. Controlling the pandemic during the SARS-CoV-2 vaccination rollout. Nat Commun 2021 Jun 16;12(1):3674 [FREE Full text] [doi: 10.1038/s41467-021-23938-8] [Medline: 34135335]
8. COVID-19 significantly impacts health services for noncommunicable diseases. World Health Organization. Geneva: World Health Organization; 2020 Jul 12. URL: https://www.who.int/news-room/detail/01-06-2020-covid-19-significantly-impacts-health-services-for-noncommunicable-diseases [accessed 2021-08-08]
9. Kocher K, Macy M. Emergency department patients in the early months of the coronavirus disease 2019 (COVID-19) pandemic—what have we learned? JAMA Health Forum 2020 Jun 09;1(6):e200705 [FREE Full text] [doi: 10.1001/jamahealthforum.2020.0705]
10. Thornton J. BMJ 2020 Apr 06;369:m1401. [doi: 10.1136/bmj.m1401] [Medline: 32253175]
11. Zylke JW, Bauchner H. Mortality and morbidity: the measure of a pandemic. JAMA 2020 Aug 04;324(5):458-459. [doi: 10.1001/jama.2020.11761] [Medline: 32609308]
12. Lapão LV, Dussault G. From policy to reality: clinical managers' views of the organizational challenges of primary care reform in Portugal. Int J Health Plann Manage 2012;27(4):295-307. [doi: 10.1002/hpm.2111] [Medline: 22648961]
13. Keesara S, Jonas A, Schulman K. Covid-19 and health care's digital revolution. N Engl J Med 2020 Jun 04;382(23):e82. [doi: 10.1056/NEJMp2005835] [Medline: 32240581]
14. Ahmed S, Sanghvi K, Yeo D. Telemedicine takes centre stage during COVID-19 pandemic. BMJ Innov 2020 Jun 12;6(4):252-254. [doi: 10.1136/bmjinnov-2020-000440]
15. Lapão LV, da Silva MM, Gregório J. Implementing an online pharmaceutical service using design science research. BMC Med Inform Decis Mak 2017 Mar 27;17(1):31 [FREE Full text] [doi: 10.1186/s12911-017-0428-2] [Medline: 28347304]
16. Reuben DB, Tinetti ME. Goal-oriented patient care--an alternative health outcomes paradigm. N Engl J Med 2012 Mar 01;366(9):777-779. [doi: 10.1056/NEJMp1113631] [Medline: 22375966]
17. Schwaber K. Agile Project Management with Scrum. Redmond, Washington: Microsoft Press; 2004.
18. Trello. Atlassian. 2021 Jun 08. URL: https://trello.com/ [accessed 2021-08-08]
19. Who's at higher risk from coronavirus (COVID-19). National Health Service. 2020 Jul 12. URL: https://www.nhs.uk/conditions/coronavirus-covid-19/people-at-higher-risk/whos-at-higher-risk-from-coronavirus/ [accessed 2021-08-08]
20. Coronavirus disease 2019 (COVID-19). Centers for Disease Control and Prevention. 2020 Jul 12. URL: https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-at-increased-risk.html [accessed 2021-08-08]
21. COVID-19 High risk groups. World Health Organization. 2020 Jul 12. URL: https://www.who.int/westernpacific/emergencies/covid-19/information/high-risk-groups [accessed 2021-08-08]

22. Maia MR, Simões A, Lapão LV. Implementing an antibiotic stewardship information system to improve hospital infection control: a co-design process. Stud Health Technol Inform 2018;247:56-60. [Medline: 29677922]
23. Peffers K, Tuunanen T, Rothenberger M, Chatterjee S. A design science research methodology for information systems research. J Manag Inf Syst 2014 Dec 08;24(3):45-77 [FREE Full text] [doi: 10.2753/MIS0742-1222240302]
24. Chowdhry SM, Mishuris RG, Mann D. Problem-oriented charting: a review. Int J Med Inform 2017 Jul;103:95-102. [doi: 10.1016/j.ijmedinf.2017.04.016] [Medline: 28551008]
25. Lapao LV, Santos M, Maia M, Pedrosa V, Seixas J, Mira da Silva M, et al. Dealing with COVID-19 barriers to care: digital platform to support and monitor chronic patients. Eur J Public Health 2020;30(Suppl 5):ckaa165.070. [doi: 10.1093/eurpub/ckaa165.070]
26. Relatório de situação. Direção-Geral da Saúde. URL: https://covid19.min-saude.pt/wp-content/uploads/2020/07/151_DGS_boletim_20200731.pdf [accessed 2021-08-12]
27. Kemp E, Trigg J, Beatty L, Christensen C, Dhillon HM, Maeder A, et al. Health literacy, digital health literacy and the implementation of digital health technologies in cancer care: the need for a strategic approach. Health Promot J Austr 2021 Feb;32 Suppl 1:104-114. [doi: 10.1002/hpja.387] [Medline: 32681656]

Abbreviations

USF: family health unit (Unidade de Saúde Familiar)

Edited by G Eysenbach, C Basch; submitted 07.09.20; peer-reviewed by S Woolley, L Barbosa, R Halkes; comments to author 02.11.20; revised version received 14.01.21; accepted 02.07.21; published 26.08.21.

Please cite as:
Lapão LV, Peyroteo M, Maia M, Seixas J, Gregório J, Mira da Silva M, Heleno B, Correia JC
Implementation of Digital Monitoring Services During the COVID-19 Pandemic for Patients With Chronic Diseases: Design Science Approach
J Med Internet Res 2021;23(8):e24181
URL: https://www.jmir.org/2021/8/e24181
doi:10.2196/24181
PMID:34313591

©Luís Velez Lapão, Mariana Peyroteo, Melanie Maia, Jorge Seixas, João Gregório, Miguel Mira da Silva, Bruno Heleno, Jorge César Correia. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Use of Iris Scanning for Biometric Recognition of Healthy Adults Participating in an Ebola Vaccine Trial in the Democratic Republic of the Congo: Mixed Methods Study

Trésor Zola Matuvanga[1,2,3], MD; Ginger Johnson[4], PhD; Ynke Larivière[2,3], MPH; Emmanuel Esanga Longomo[5], MD; Junior Matangila[1], PhD; Vivi Maketa[1], PhD; Bruno Lapika[6], PhD; Patrick Mitashi[1], PhD; Paula Mc Kenna[7], MSc; Jessie De Bie[2,3], PhD; Jean-Pierre Van Geertruyden[2], PhD; Pierre Van Damme[3], PhD; Hypolite Muhindo Mavoko[1], PhD

[1]Department of Tropical Medicine, University of Kinshasa, Kinshasa, the Democratic Republic of the Congo
[2]Global Health Institute, University of Antwerp, Antwerp, Belgium
[3]Centre for the Evaluation of Vaccination, University of Antwerp, Antwerp, Belgium
[4]Department of Public Health, Institute of Tropical Medicine, Antwerp, Belgium
[5]Division Provinciale de la Santé de la Province de la Tshuapa, Boende, the Democratic Republic of the Congo
[6]Department of Anthropology, University of Kinshasa, Kinshasa, the Democratic Republic of the Congo
[7]Janssen Pharmaceutica NV, Beerse, Belgium

Corresponding Author:
Trésor Zola Matuvanga, MD
Department of Tropical Medicine
University of Kinshasa
Avenue Université numéro 1, Commune de Lemba
Kinshasa
the Democratic Republic of the Congo
Phone: 243 810046306
Email: zola.matuvanga@unikin.ac.cd

Abstract

Background: A partnership between the University of Antwerp and the University of Kinshasa implemented the EBOVAC3 clinical trial with an Ebola vaccine regimen administered to health care provider participants in Tshuapa Province, Democratic Republic of the Congo. This randomized controlled trial was part of an Ebola outbreak preparedness initiative financed through Innovative Medicines Initiative-European Union. The EBOVAC3 clinical trial used iris scan technology to identify all health care provider participants enrolled in the vaccine trial, to ensure that the right participant received the right vaccine at the right visit.

Objective: We aimed to assess the acceptability, accuracy, and feasibility of iris scan technology as an identification method within a population of health care provider participants in a vaccine trial in a remote setting.

Methods: We used a mixed methods study. The acceptability was assessed prior to the trial through 12 focus group discussions (FGDs) and was assessed at enrollment. Feasibility and accuracy research was conducted using a longitudinal trial study design, where iris scanning was compared with the unique study ID card to identify health care provider participants at enrollment and at their follow-up visits.

Results: During the FGDs, health care provider participants were mainly concerned about the iris scan technology causing physical problems to their eyes or exposing them to spiritual problems through sorcery. However, 99% (85/86; 95% CI 97.1-100.0) of health care provider participants in the FGDs agreed to be identified by the iris scan. Also, at enrollment, 99.0% (692/699; 95% CI 98.2-99.7) of health care provider participants accepted to be identified by iris scan. Iris scan technology correctly identified 93.1% (636/683; 95% CI 91.2-95.0) of the participants returning for scheduled follow-up visits. The iris scanning operation lasted 2 minutes or less for 96.0% (656/683; 95% CI 94.6-97.5), and 1 attempt was enough to identify the majority of study participants (475/683, 69.5%; 95% CI 66.1-73.0).

Conclusions: Iris scans are highly acceptable as an identification tool in a clinical trial for health care provider participants in a remote setting. Its operationalization during the trial demonstrated a high level of accuracy that can reliably identify individuals.

Iris scanning is found to be feasible in clinical trials but requires a trained operator to reduce the duration and the number of attempts to identify a participant.

Trial Registration: ClinicalTrials.gov NCT04186000; https://clinicaltrials.gov/ct2/show/NCT04186000

(J Med Internet Res 2021;23(8):e28573) doi:10.2196/28573

KEYWORDS

biometric identification; iris recognition; vaccine trial; participants' visits; acceptability; feasibility; Democratic Republic of the Congo; mixed methods; Ebola

Introduction

Identification and recognition of study participants in a clinical trial—during the process of recruitment and during follow-up visits—is a growing issue [1]. Conventional methods for the recognition of participants in health facilities may include patient name, date of birth, government identity card with photo, and phone number [2-5]. However, these methods are not always reliable or accurate [5]. For example, identity cards can be stolen or forgotten, and there is a risk of assigning a participant's ID (intentionally or unintentionally) to another participant during a study visit. Some participants may share their ID card number with a family member with a similar physical resemblance, if they are unable or unwilling to keep to their appointment time. In clinical trials, efficacy and safety data such as (serious) adverse events, are repeatedly assessed through anamneses, physical examinations, and biological samples during different visits, possibly, over a long period of time. Thus, participant enrollment and identification are essential steps to ensure that all data collected are unique and that neither the participant nor the visit has been misidentified [1,2]. A biometric identification method coupled with a unique participant ID number could mitigate the occurrence of mistakes made using conventional methods during initial and follow-up clinical trial visits [3].

Biometric technology confirms the physical presence of the person by assessing unique physical or behavioral characteristics that cannot be borrowed, stolen, or forgotten. Such technology uses matching algorithms or artificial intelligence for identifying the particular feature [3,6,7]. A number of biometric identifiers, including physical traits (eg, fingerprint; face; palm; cornea; iris; thermogram of the body, face or ear; and DNA) or behavioral traits (eg, signature, voice, typing dynamic, smell, and walk pattern) have demonstrated technical feasibility in various studies [6-10]. Biometric identification systems have many advantages over more conventional methods of identification, such as easier fraud detection and more accuracy for face recognition than photographs. Therefore, biometric identification is increasingly used worldwide in various fields to recognize individuals and secure their data (eg, during elections, at airports, or for criminal detection) [6].

Irises are an ideal part of the body for biometric identification. The iris is flat and has a fine texture and geometric configuration determined randomly upon embryogenesis [3,10,11]. It is a unique, permanent, and universal “biometric signature” present throughout a person’s lifespan, which is covered by a highly transparent and sensitive membrane that makes it distinctive from other biometric methods [1,2]. A human iris is always stable, irrespective of age [10]. This is in contrast to the fingerprint structure—the most widespread biometric method of identification—that varies during childhood and only becomes stable after many years [5,9]. Fingerprinting also carries additional risks such as spreading some infectious diseases, since it requires the participant (and, sometimes, the operator) to come in physical contact with the fingerprint device. Identical (ie, monozygotic) twins were found to have higher similarities of fingerprint patterns compared with nonidentical twins [4,5]. Iris scanning is feasible under most circumstances, as it can be carried out from anywhere between 10 cm to several meters away from the eye, and results are generally available within 30 seconds [7,12]. Even genetically similar people have entirely independent irises; thus, iris scanning recognition avoids misidentification of identical twins [3-5]. However, iris recognition may be challenging for people who suffer from diabetes or any other iris disease [4,5]. Moreover, the accuracy of the scanning devices can be affected by unusual light effects, in comparison with fingerprinting [2,5].

Iris scans may offer one of the most secure strategies of authentication and recognition in clinical trials [7]. Iris-based biometric systems have demonstrated a promising performance during the process of recognition, with an average time (during initial clinical trial enrollment) of less than 2 minutes and a sensitivity rate of at least 86% [8,11,12]. In Kenya, an iris scan sensitivity or accuracy rate of 95% was found in HIV and tuberculosis patients during routine hospital consultations [8]. This was better than fingerprint biometric recognition found in Ghana (68.7%) or in Uganda (75.5%) [11,12]. Thus, use of iris scan technology can substantially reduce the possibility for fraud and abuse within a clinical trial [3,9,10]. Lastly, it has a high acceptance rate, with very low false-match and rejection rates [1,2].

Despite its attractive design features, there is little information available about the acceptability of iris scan technology for the general public, especially, information on how it varies across and within countries. Acceptability within a population may depend on many factors such as positive perception, confidence, and constraints presented against the use of iris scans. For example, in one of the few studies available, a survey was conducted in Australia on the willingness of the general population to use biometric security technologies; it found that 61% of the population would accept fingerprints, whereas only 41% would accept iris scan recognition [6]. In California, 72% of participants preferred an identification by fingerprint [9]. A remarkable acceptability rate of iris scanning itself (98.9%) was noted in a survey on an identification system of routine clinic services in Kenya [8].

As part of an ongoing Ebola vaccine clinical trial (EBOVAC3, study protocol number VAC52150EBL2007, clinicaltrials.gov identifier NCT04186000) [13], we assessed acceptability, accuracy, and feasibility of iris scan technology as a biometric identification method within a population of health care provider participants in a remote setting.

Methods

Study Design

A mixed methods study design assessed the acceptability, accuracy, and feasibility of the iris scan as a biometric identification tool in the Ebola vaccine trial in Boende, Tshuapa province, Democratic Republic of the Congo (DRC). Acceptability was assessed through focus group discussions (FGDs) with volunteering health care provider participants and via a survey with a structured questionnaire. Feasibility and accuracy research was conducted using a longitudinal study design, where iris scanning was used to uniquely identify health care provider participants at enrollment and at their follow-up visits in the clinical trial. Accuracy and feasibility studies were conducted from December 2019 to April 2020, from the second participant visit (day 57) until the third participant visit to the study site (day 78) (Multimedia Appendix 1).

Participants and Recruitment Procedures

For the qualitative acceptability assessment, study participants were selected using purposive, nonprobability sampling. A total of 86 participants were enrolled in 12 focus FGDs (Multimedia Appendix 1). For the FGDs, we selected key informants from the following stakeholder groups: nurses, community health workers, laboratory technicians, medical doctors, first-aid officers, birth attendants, and hospital cleaners. All recruited health care provider participants worked at the reference hospital or health centers within the Boende District, with the exception of nurses, who worked in health centers throughout Tshuapa Province. Research activities occurred at 5 sites, all located in the Boende Health Zone: Boende General Hospital (ie, Hôpital Général de Référence de Boende), Boende Catholic Mission, N'sele Health Center (ie, Centre de Santé Boende II N'sele), Motema Mosantu Health Center (ie, Centre de Santé Motema Mosantu), and Communauté des Disciples du Christ au Congo Health Center (ie, Centre de Santé CDCC).

For the quantitative study component (assessing acceptability, accuracy, and feasibility), all health care provider participants enrolled in the clinical study (N=699) were included. All participants were health care provider participants working in the Boende Health District. Their workstations were located between 0 kilometers and 50 kilometers away from Boende General Hospital.

Ethical Approval

Research was conducted in line with the prevailing ethical principles of socio-behavioral studies with human populations to protect the rights and welfare of all participants. Permission to undertake the acceptability (qualitative) study was granted by the DRC National Ethics Committee for Health (reference 93/CNES/BN/PMMF/2019), the Institute of Tropical Medicine, Belgium (reference 1293/19), and the University of Antwerp, Belgium (reference 19/14/188). Permission for the accuracy and feasibility (quantitative) study, collected during the course of the ongoing clinical trial, was granted by the DRC Ethical National Committee (reference 137/CNES/BN/PMMF/2019).

Data Collection and Informed Consent

Pretrial Study

Based on a literature review, a topic guide was developed highlighting potential key issues with regard to the acceptance of new technologies among health care providers in DRC. This review formed the basis for the design of the FGD tool, which included questions and probes focusing on the background of health care provider participants and their role in the community, their acceptance of new technologies and communication strategies, and their recommendations for appropriate identification and communication tools with trial participants (Multimedia Appendix 1). University of Antwerp and University of Kinshasa (UNIKIN) team members reviewed and refined the research tools prior to their finalization and implementation. Specific questions and probes were reviewed and refined during the research period, in light of arising themes (eg, an ongoing Monkeypox vaccine trial in Tshuapa at the time of data collection) [14].

Key topics were addressed in each discussion, to allow for generalization of themes across participant groups. The research was deliberately designed to facilitate input from multiple health care provider participant stakeholders in a step-wise manner, so that issues raised by one group of participants were also discussed with other participant groups to assist with triangulation of data. At the start of each discussion, it was made clear to all potential participants that their involvement was optional and voluntary. The study's consent form was presented and explained in detail, and all participants' questions were answered prior to beginning data collection. Informed consent was given verbally. All FGDs were conducted in either French or Lingala, depending on the linguistic preferences of participants. FGDs lasted for approximately 60-80 minutes. Audio recordings were made, along with field notes, which served as the basis for a thematic analysis of data. Concurrent to FGDs, acceptability was (quantitatively) defined as the number of participants agreeing to iris scanning as a proportion of all the individuals approached. Reasons for declining iris scanning were elicited from participants.

Intratrial Study

Accuracy was measured by the rate of successful recognition of study participants (percentage of participants recognized by the iris scan) during the participants' third visit (day 78). This was achieved by cross-referencing the output of the iris scan with the clinical trial identity card of each participant, to make sure that it was indeed the correct study participants returning on their corresponding scheduled visit dates. We considered it a wrong match when registered participants returned for their next visit and the system gave details of more than one possible identification record. Feasibility was measured by how long (ie, duration of operation based on ranges ≤1 min, 1 min 1 s-1 min 30 s, 1 min 31 s-2 min, 2 min 1 s-2 min 30 s, or ≥2 min 30 s; Table 1, Table 2, and Table 3) the iris scanning device took to recognize each study participant in the EBL2007 Ebola vaccine trial during the second and the third visit (ie, day 57 and day 78, respectively) and the number of scanning attempts that were required by the iris scan operator or by the iris scan devices (ie, tablet, scanner, server, and Wi-Fi connection between server and tablet) during these same visits. The duration of operation included the time for the biometric tablet to capture the iris image, identity photo, and demographics and the time it took to link these data with the local server. An assessment of time to recognize each study participant at their third visit was recorded by the operator. It is important to note that a problem was encountered during the first study visit (day 1), where all vaccinated participants who received their first vaccine dose on that day should have had their demographic information and iris scan recognition registered on the server. However, these data were lost due to a manual error that occurred when attempting to save all of the data collected for this visit, resulting in the loss of participant demographic and biometric data. This error was corrected during the second visit (day 57), when all participant data were re-entered (Multimedia Appendix 1).

Table 1. Duration of the initial iris scan process in the EBL2007 clinical trial in Boende, Tshuapa Province, Democratic Republic of the Congo (N=683).

Duration of iris scanning operation to record subjects	Frequency, n (%)	95% CI	Cumulative percentage, %
0 s-1 min	280 (41.0)	37.3-44.7	41.0
1 min 1 s-1 min 30 s	332 (48.6)	44.9-52.4	89.6
1 min 31 s-2 min	44 (6.5)	4.6-8.3	96.1
2 min 1 s-2 min 30 s	24 (3.5)	2.1-4.9	99.6
≥2 min 30 s	3 (0.4)	0.0-0.9	100

Table 2. Iris scan attempts at the second visit in the EBL2007 clinical trial (N=683).

Number of iris scanning attempts	Frequency, n (%)	95% CI	Cumulative percentage, %
Once	475 (69.6)	66.1-73.0	69.5
Twice	149 (21.8)	18.7-24.9	91.4
Three times	59 (8.6)	6.5-10.8	100

Table 3. Duration of the iris scan process during the third visit in the EBL2007 clinical trial (N=683).

Duration of the operation at the third visit	Frequency, n (%)	95% CI	Cumulative percent, %
0 s-1 min	665 (97.4)	96.1-98.6	97.4
1 min 1 s-1 min 30 s	1 (0.1)	0.0-0.4	97.5
1 min 31 s-2 min	9 (1.3)	0.5-2.1	98.8
2 min 1 s-2 min 30 s	5 (0.7)	0.1-1.3	99.5
≥2 min 30 s	3 (0.5)	0.0-1.0	100

Equipment and Procedures

The iris scan operator, a trained and authorized study staff member, used an iris camera (Iritech, Irishield Monocular Fairfax, VA 22030, United States), and a tablet (Samsung Tab Active 2, Suwon, South Korea) connected via Wi-Fi to a local ruggedized server (Cincoze DX-1100, New Taipei City, Taiwan) located approximately 10 meters from his physical location. An external hard drive for backing up the iris scanning database was located nearby as well. A biometric user interface running on the Samsung Tab Active 2 was designed by Janssen Pharmaceutica NV Beerse, Belgium. In addition to the iris scan, the operator captured demographic data on the biometric tablet, such as gender, year of birth, participant ID, passport photo, contact telephone number, date, and time stamps of the iris scan. Activities performed on the tablet were captured in an audit trail with date and time stamps. The biometric tablet allowed the operator to assess whether administration of the second vaccine dose (administered on day 57) as well as the blood collection during the third visit (day 78) were in the predefined visit windows or not. To capture the irises, the operator stood in front of the study participant and held a camera in their right hand and a tablet in their left hand. Participants were seated so that their head and body were vertically aligned. The distance between the iris and the camera ranged from 3-10 cm.

Data Analysis

Pretrial Study

At the conclusion of the research activities, the lead qualitative researcher, GJ, a medical anthropologist, transcribed notes, alongside rereviewing audio files to compile data for review and verification. Notes were typed in either English or French. A preliminary analysis of qualitative data was conducted throughout the data-collection process. The lead researcher, GJ, was responsible for all thematic analysis of qualitative data. Dominant themes were identified through the systematic review of FGD audio and transcribed notes. The occurrence and reoccurrence of salient concepts were labelled throughout, and emerging trends were critically analyzed according to the research objectives and topic guide. An appointed research member, TZ, was additionally responsible for maintaining the quantitative survey database with the acceptance rate of iris scanning at the end of discussions.

Intratrial Study

Data with regard to the accuracy and feasibility were collected in an Excel spreadsheet. The dataset was checked for any inconsistencies such as duplicates and then processed using Excel to synthesize the results in terms of proportions.

Results

Pretrial Study: Acceptability and Concerns About the Iris Scan

Data collection and in-country fieldwork were conducted in April 2019. Overall, acceptance (85/86, 99%; 95% CI 97.1-100.0) of the iris scan technology was widespread. As stated by one nurse:

> *For me, I accept [iris scanning], because I know it is a process that is being used to cast away Ebola, and we want this disease to leave.*

However, another research participant, also a nurse, refused to have her picture taken while consenting to have her iris scanned, on the justification that the picture may cause problems with her church superiors.

FGD participants voiced some primary concerns about the iris scan. The concern that the iris scan may cause physical problems to their eyes was widespread, across all stakeholder groups and education levels. As stated by one community health worker:

> *We are agreeing with what you say, but we are afraid with the use of the eye scan, because we fear it may cause problems with our eyes.*

Similarly, one birth attendant stated:

> *We are asking because the eyes are the life of the people, so, after using the eye scan, will there be some problems, for us, with our eyes?*

Participants often associated the extended duration of some scans as harmful to their eyes due to the light emitted by the scanner. The research team often heard participants asking:

> *Will the scanner disturb the eyes with the [light] rays, in relation to the duration?*

Participants also asked:

> *Are you sure that this scan will not hurt our eyes?*

The pretrial acceptability study, therefore, noted that there was a higher risk for the participants enrolled in the trial to link any vision loss to the iris scan. In fact, even if a participant consented to an eye scan at the time of the vaccination as indicated by the quantitative survey (ie "we are agreeing with you"), any problems pertaining to eyes (through naturally occurring means) could later be associated with the iris scan. This is illustrated by the following exchange with a laboratory technician:

> *We are using the microscope, and we are suffering from our eyes because of looking through the microscope, so maybe we will have a problem in the long-term with our eyes...our first thought will be that the technology caused this problem, so this is why we need a very good explanation, so that we know it is not the technology that is causing the problem.*

A second concern was wondering, will the "iris scan...expose me to spiritual problems through sorcery?" This was discussed by most stakeholder groups as primarily a problem for "those who are not learned" or those who belong to churches that reject vaccination (ie, "some churches here that are proving to the population that they should not receive a vaccination"). For example, discussions with doctors, nurses, and laboratory technicians regarding persons who may be concerned about the potential of the technology to open them up to witchcraft, often started with the phrase:

> *For us, there is no problem [with the iris scan], but other people will need to be sensitized to accept…[T]he education level of the population is very low. If you use the eye scan, they may think you are trying to make trouble through their eyes.*

By using the phrase "us," health care provider participants are referencing persons such as themselves who are well-educated health professionals. This sentiment did not often extend to "other" stakeholder groups (eg, community health workers) with a lower-level of education. This comment from a laboratory technician is illustrative:

> *By using the eye scan, many people will be having a bad [thought] that the eye scan will cause trouble with the eyes, and people will run away,...because, if you are using the eye scanner, they will think you are putting something into their eyes.*

The pretrial acceptability study, therefore, noted that regardless of whether or not the health care provider participants who are enrolled in the trial harbor suspicions about the technology with regard to witchcraft, they are embedded in the larger cultural and religious communities of Tshuapa, who are likely to have such concerns. As such, trial organizers should be aware of, and have a communication plan prepared for, the potential myths and rumors that may manifest in Tshuapa, which associate the iris scan with the evil intentions of witchcraft.

Three types of identification were familiar and considered to adequately identify vaccine recipients: ID cards (containing name, address, phone number, etc), thumbprints or fingerprints, and facial photographs. An ID card as a method of identification was used in the Monkeypox vaccine trial, which was still occurring in the area while the qualitative pre-EBOVAC3 trial activities were ongoing. This was used to identify the trial enrollees. Several doctors familiar with the Monkeypox vaccine trial felt it may be confusing for some EBOVAC3 participants to be requested to have their eye scanned as a method of identification, given their familiarity with a different method as established by the recent Monkeypox vaccine trial. Participants also felt strongly that the use of a facial photograph by itself (without scanning both eyes) was a sufficient method to identify individual persons. As stated by one nurse:

> *If the iris scan is just a picture of the eye, why not just take a picture of the person? This is also a positive way to identify them, which does not take so much time.*

A community health worker similarly stated:

> *I can change my clothes, but I can't change my face.*

In general, participants were confused as to why 3 pictures—1 of their face and 1 of each eye—were necessary as a method of identification. While the iris scan technology was not rejected by most participants, many favored the use of ID cards plus facial photographs as a positive method of identification. Preference for photos of their face rather than a scan of each eye was due to considering photos as less invasive and less time-consuming but equally positive as a way to identify an individual.

Intratrial Study: Acceptance, Accuracy, and Feasibility of Iris Scan Identification Technology

It was noted that, of 699 participants enrolled, 99.0% (692/699; 95% CI 98.2-99.7) had given consent to be identified by the iris scan technology. Thus, 7/699 (1.0%; 95% CI 0.3-1.7) of the participants refused. Various reasons were given for refusing, but most of them argued about the fear of alterations to their visual acuity over time (Table 4). Among the participants who agreed to have their iris scanned, 0.9% (6/692; 95% CI 0.2-1.6) did not return for the second or third visits. In addition, iris scan data of 0.4% (3/692; 95% CI 0.0-0.9) of the participants were not properly entered in the database at the second visit due to inattention by the iris scan operator during the registration process of inputting data in the server. As a result, the quantitative survey conducted for the accuracy and feasibility study was only possible for 683/692 (98.7%) of participants who agreed to be identified by an iris scan during their initial clinical trial visits.

Capturing a successful and quick iris scan is a process requiring both a participant who is willing to follow operator instructions (eg, face forward, chin down, etc) and a skilled operator capable of balancing the tablet in one hand while successfully locating the iris with the scanning device in the other hand. It often took more than one attempt to receive feedback on the tablet screen that a participant's irises were correctly scanned in the iris scan server (Table 2).

During the process of rerecording each participant's iris scans for both the first and second visit of the clinical trial, the duration of the operation ranged from 1 minute 1 second to 1 minute 30 seconds, for the majority of study participants (332/683, 48.6%; Table 1). Capturing a successful image of the iris often took several seconds and required multiple manipulations of participants' face and body by the iris scan operator, in order to obtain a successful reading (Table 1). This concern seemed to exacerbate participant conclusions that the eye scan was taking too long and potentially causing long-term damage to their eyes. The process of recording study participants by scanning their irises, capturing a photo, and entering their demographic data into the tablet lasted 2 minutes or less for 96.0% (656/683; 95% CI 94.6-97.5) of participants. At the third visit, it took less than 1 minute for 97.4% (665/683; 95% CI 96.1-98.6) of participants to be authenticated. Overall accuracy of the iris scan, calculated by the percentage of successful iris scanning recognitions on the third visit, was 93.1% (636/683; 95% CI 91.2-95.0).

Table 4. Participants in the EBL2007 clinical trial who refused to be identified by an iris scan.

Gender	Iris scan performed? (yes or no)	Reason for denial of iris scan
Male	No	Fears the scanner will cause defective vision in the future
Male	No	Deteriorated vision prior to enrollment in the clinical trial and fears the scanner will further damage his eyes
Female	No	Fear of the iris scanning tools and devices
Male	No	Fears the scanner will cause defective vision in the future
Female	No	Fear of the iris scanning tools and devices; fears the scanner will cause defective vision in the future
Male	No	Fear of the iris scanning tools and devices
Male	No	Fears the scanner will cause defective vision in the future

Discussion

Principal Findings

In general, iris scanning as a biometric technology for identifying participants in a clinical trial was acceptable, feasible, and accurate. A high acceptability rate (99.1% pretrial; 99.0% intratrial) of biometric identification via iris scanning was noted among the health care provider participants. This remarkable rate of acceptance was similar to the one found in the quantitative survey conducted prior to the implementation of this technology for the clinical trial.

Results from the quantitative survey should be interpreted with care, as health care provider participants may not be representative of the general population of Tshuapa province or elsewhere. The qualitative data presented here describe a more nuanced picture of technology acceptance (eg, concerns over physical or spiritual problems from the iris scan) than the reported quantitative survey results alone. Prior to starting the clinical trial, the quantitative survey conducted among potential trial participants found that less than 1.0% of them refused the irises scans and preferred other identification methods such as simply capturing a photo on a participant's card or registering fingerprints. This low refusal rate was confirmed during the implementation of the trial. This is an illustration of the need to anticipate risk perceptions in a community of potential clinical trial participants via a prior acceptability study, in order to determine beforehand whether or not that community is ready to use an innovative biometric identification technology. To our knowledge, this is the first study in sub-Saharan Africa demonstrating the use of iris recognition in a clinical trial involving an adult population. Our high acceptability (99.1% pretrial; 99.0% intratrial) is comparable with other observational studies using iris scans in Kenya and Brazil [8,15]. This is likely because potential participants were already briefed on the value of using iris scans in the trial prior to the start of the study through a previous workshop. The FGDs conducted during the pretrial qualitative study helped clinical trial investigators "empty all pockets of fear" with regard to the use of this innovative technology. Furthermore, demonstrations of the functionality of this tool, the explanations given during the qualitative survey, and the ability to explore participants' potential fears and concerns about the technology through FGDs likely had an influence on the willingness of health care provider participants to accept the iris scan as an identification technique.

Various reasons were given by trial participants for potential refusals of the iris scan, but most were fearful that their visual acuity would be altered over time. Fears associated with a new and unknown technology needed to be overcome, not only by volunteers but also by the iris scan operator, who struggled at the start of the trial with using a new technology and making sure that all participant details were recorded accurately and quickly (to limit participant fears). That is, while implementing the iris scan in the clinical trial, several issues did arise with regard to some extended wait times for receiving the feedback that a good-quality iris scan stamp was properly recorded, before the ability to enter participants' other demographic details. In addition, the use of the tablet to instantaneously capture an image of the participant's face, prior to proceeding to scan the eye, caused participants to conclude that the eye scan was taking too long. Capturing an image of the participant's face was always quickly and immediately successful, without any special posturing by the participant. However, capturing a successful image of the iris often took several seconds and required multiple manipulations of their face and body by the iris scan operator in order to obtain a successful reading. This sometimes caused whispering and fatigue in the queue of health care provider participants, who were often impatient about waiting in line for their turn, especially, when the iris scanner took 2 minutes or more to successfully capture 1 iris scan. The longer duration at the second visit was, however, due to data re-entering that had to be performed for both visit 1 and visit 2. That would likely not have been the case if the manual error had not occurred after the day-1 visit, which could have saved time and speculations from the participants. During the third visit, things were easier for the iris scan operator, as only one scan of an iris was enough for the system to provide the picture and the appointment window of the participant. Continuous practice by the operator is, therefore, important for the success of using this technology.

Finally, quantitative research demonstrated that iris scanning technology can be used effectively in clinical trials in resource-poor countries. An accuracy rate of 93.1% in this study is better, compared with the 85% accuracy reported in Brazil [15]. However, the accuracy rate in Kenya was even higher, at 95% [8]. With these appreciable accuracy rates, iris scan technology demonstrates the importance of scaling it up in the future, for widespread use in clinical trials and for the automation of subjects' identification processes. The time duration required to capture the iris scan and other related information of health care provider participants at enrollment is similar to that reported in Brazil (less than 2 minutes) [15]. This time is shorter than the average of 4 minutes reported in Kenya [8]. It is understood that this time depends on the amount of information needed for each person included in the study, and that this is a factor that influences the recording time at recruitment. It should also be pointed out that the accuracy rate of 93.1% may have been underestimated, given that a failed recognition was scored even when the correct matching profile was presented along with other possible matching profiles upon completion of the iris-recognition process. The training of scanning operators, for the steps to take if during the matching process multiple profiles are offered, is more important here and, to a lesser degree, an issue of practice.

Some weaknesses were found that could be attributed either to the operator or to the iris scan system. With regard to the operators, if they did not scan both irises with equal precision (after each iris scan, the biometric tool showed the iris scan precision with color codes green [high], orange [medium], or red [low precision]), the biometrical tool sometimes provided several possible participant matches as an output. Based on a photograph entered at the beginning of the trial, the operator could then select the correct participant. Similarly, an issue sometimes occurred when the operators did not correctly enter the study ID for the participant in front of them, which is the basis for the pop-ups of the participant information on the tablet. A loss of information is also possible, as was the case after the first visit of this study, if the operator or the site does not pay attention to the standard operating procedure of the system (eg, how to save the recorded information). This would constitute a deadlock to identify participants for the future visits. In the event of a possible false match or a correct match, the identity of each participant was to be double-checked using the profile picture and biographical data also entered into the biometric tool, as well as with a participant ID card. Yet, cross-referencing outputs of a biometric tool to an ID card may present a risk in underestimating accuracy of the biometric tool. In fact, the use of the ID card as a reference, which can be tampered with, can compromise the benefit of the biometric tool in detecting fraud [4].

Nonetheless, it is worth mentioning that during subsequent visits, the iris scan allowed detection of some cases of fraud attempts. For example, some people not enrolled in the trial tried to come on a scheduled visit to replace a relative. In addition, a few study participants attempted to falsify their ID numbers in order to change the study's activity schedule (ie, vaccination or blood sample collection). In these cases, the iris scanning system was able to catch the attempted fraud. Moreover, the Ebola vaccine trial has quite a long follow-up period. This highlights the relevance of using this technology to correctly identify the clinical trial participants, to make sure, for example, that a blood sample is collected from participants who actually received the study intervention and not from their relatives.

It seems important to consider a qualitative study that measures trial participants' perceptions after identification by an iris scan, as these perceptions can add value. This would provide a better understanding of the contours of the level of acceptability among health care provider participants in a vaccine trial.

During the assessment of acceptability, accuracy, and feasibility of the iris scan system, nearly no technical issues were encountered. The equipment that was used had the advantage of lacking dependency on internet connection, as the device was connected to the server via a local Wi-Fi connection. In a few occasions where a technical problem occurred, it was troubleshooted automatically by rebooting or bringing the tablet closer to the server. Both the tablet and the server needed a power supply. Hence, its implementation in a remote area should take this into account beforehand. In the Ebola vaccine trial in Boende, a generator was running permanently onsite, and uninterrupted power supplies were available as back-ups.

Conclusions

Identification through iris scanning is an innovative technology that was found to be acceptable, accurate, and feasible with health care provider participants in a remote setting. This biotechnical tool takes little additional time, can automate the process of identifying subjects in a clinical study, and can quickly recall relevant information in relation to trial appointments. Thus, it helps to guarantee the quality of data. Sensitization of investigators and potential study participants and their communities is a necessary prerequisite to successfully introduce this promising technology in trials conducted in low- and middle-incomes countries. We hope that this paper will, therefore, spark the idea of proposing further explorations in the field of biometric identification technology. Then, solutions could be found for the difficulties encountered here, to further leverage performance of the iris scan as a fast and reliable biometric method to implement in clinical trials.

Acknowledgments

Particular gratitude is extended to the health care workers and volunteers who have (and continue to) participate in this study. We would also thank the following individuals and institutions for their support and inputs: Dr Jean-Paul Matela from the Division Provinciale de la Santé (DPS), Boende, DRC; Gwen Lemey and Paul Peter Vermeiren from the University of Antwerp, Antwerp, Belgium; and Professor Dr Koen Peeters from the Institute of Tropical Medicine, Antwerp, Belgium. We thank the Iris Scan technology development team of Johnson & Johnson Global Public Health for providing this technology, delivering training, and supporting our teams for the implementation, especially, Annik Willems, Serge Masyn, and Romain Rutten.

Within the Innovative Medicines Initiative-European Union (IMI-EU) project EBOVAC3, the University of Antwerp, along with University of Kinshasa (UNIKIN) partners, is conducting a phase-2 vaccine trial with a candidate Ebola vaccine in Boende in a healthy population of health care providers (study protocol number VAC52150, clinicaltrials.gov identifier NCT04186000). Janssen Pharmaceutica (J&J) provided iris scanning equipment to be evaluated for use in this trial. This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking, under grant number 800176. Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation program and the European Federation of Pharmaceutical Industries and Associations.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Completed and planned events of EBL2007 clinical trial, focus group discussion stakeholder groups, and research tool.
[PPTX File , 162 KB - jmir_v23i8e28573_app1.pptx]

References

1. National Research Council of the National Academies. In: Pato JN, Millett LI, editors. Biometric Recognition: Challenges and Opportunities. Washington, DC: National Academies Press; 2010.
2. Aslam TM, Tan SZ, Dhillon B. Iris recognition in the presence of ocular disease. J R Soc Interface 2009 May 06;6(34):489-493 [FREE Full text] [doi: 10.1098/rsif.2008.0530] [Medline: 19324690]
3. Lu L, Mao J, Wang W, Ding G, Zhang Z. A Study of Personal Recognition Method Based on EMG Signal. IEEE Trans Biomed Circuits Syst 2020 Aug;14(4):681-691. [doi: 10.1109/TBCAS.2020.3005148] [Medline: 32746348]
4. Masyn S, Vuchelen A, Santermans E, Rasschaert F, Bangura A, Parys W, et al. Overcoming the challenges of iris scanning to identify minors (1-4 years) in the real-world setting. BMC Res Notes 2019 Jul 22;12(1):448 [FREE Full text] [doi: 10.1186/s13104-019-4485-8] [Medline: 31331369]
5. Tao X, Chen X, Yang X, Tian J. Fingerprint recognition with identical twin fingerprints. PLoS One 2012;7(4):e35704 [FREE Full text] [doi: 10.1371/journal.pone.0035704] [Medline: 22558204]

6. Binder S, Iannone A, Leibner C. Biometric technology in "no-gate border crossing solutions" under consideration of privacy, ethical, regulatory and social acceptance. Multimed Tools Appl 2020 Dec 29:1-14 [FREE Full text] [doi: 10.1007/s11042-020-10266-0] [Medline: 33390767]
7. Gassman JJ, Owen WW, Kuntz TE, Martin JP, Amoroso WP. Data quality assurance, monitoring, and reporting. Control Clin Trials 1995 Apr;16(2 Suppl):104S-136S. [doi: 10.1016/0197-2456(94)00095-k] [Medline: 7789140]
8. Anne N, Dunbar MD, Abuna F, Simpson P, Macharia P, Betz B, et al. Feasibility and acceptability of an iris biometric system for unique patient identification in routine HIV services in Kenya. Int J Med Inform 2020 Jan;133:104006 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.104006] [Medline: 31706230]
9. Cohen JK, Flynn R, Bolan R, Klausner JD. Acceptability of fingerprint scanning for personal identification among patients seeking HIV/STI-related services, Los Angeles, 2011. J Acquir Immune Defic Syndr 2012 Dec 01;61(4):e59-e60. [doi: 10.1097/QAI.0b013e318271481a] [Medline: 23138730]
10. Unar J, Seng WC, Abbasi A. A review of biometric technology along with trends and prospects. Pattern Recognition 2014 Aug;47(8):2673-2688 [FREE Full text] [doi: 10.1016/j.patcog.2014.01.016]
11. Odei-Lartey EO, Boateng D, Danso S, Kwarteng A, Abokyi L, Amenga-Etego S, et al. The application of a biometric identification technique for linking community and hospital data in rural Ghana. Glob Health Action 2016;9:29854 [FREE Full text] [Medline: 26993473]
12. White EB, Meyer AJ, Ggita JM, Babirye D, Mark D, Ayakaka I, et al. Feasibility, Acceptability, and Adoption of Digital Fingerprinting During Contact Investigation for Tuberculosis in Kampala, Uganda: A Parallel-Convergent Mixed-Methods Analysis. J Med Internet Res 2018 Nov 15;20(11):e11541 [FREE Full text] [doi: 10.2196/11541] [Medline: 30442637]
13. Study to Evaluate the Immunogenicity and Safety of a Heterologous Vaccine Regimen Against Ebola (EBOVAC3DRC). URL: https://clinicaltrials.gov/ct2/show/NCT04186000 [accessed 2019-12-04]
14. IMVAMUNE® Smallpox Vaccine in Adult Healthcare Personnel at Risk for Monkeypox in the Democratic Republic of the Congo. URL: https://clinicaltrials.gov/ct2/show/NCT02977715 [accessed 2016-11-30]
15. Corby PM, Schleyer T, Spallek H, Hart TC, Weyant RJ, Corby AL, et al. Using biometrics for participant identification in a research study: a case report. J Am Med Inform Assoc 2006;13(2):233-235 [FREE Full text] [doi: 10.1197/jamia.M1793] [Medline: 16357359]

Abbreviations

DRC: Democratic Republic of the Congo
FGD: focus group discussion

Edited by R Kukafka; submitted 06.03.21; peer-reviewed by J Marconato, T Capote, P Macharia; comments to author 27.04.21; revised version received 11.05.21; accepted 30.06.21; published 09.08.21.

Please cite as:
Zola Matuvanga T, Johnson G, Larivière Y, Esanga Longomo E, Matangila J, Maketa V, Lapika B, Mitashi P, Mc Kenna P, De Bie J, Van Geertruyden JP, Van Damme P, Muhindo Mavoko H
Use of Iris Scanning for Biometric Recognition of Healthy Adults Participating in an Ebola Vaccine Trial in the Democratic Republic of the Congo: Mixed Methods Study
J Med Internet Res 2021;23(8):e28573
URL: https://www.jmir.org/2021/8/e28573
doi:10.2196/28573
PMID:34378545

©Trésor Zola Matuvanga, Ginger Johnson, Ynke Larivière, Emmanuel Esanga Longomo, Junior Matangila, Vivi Maketa, Bruno Lapika, Patrick Mitashi, Paula Mc Kenna, Jessie De Bie, Jean-Pierre Van Geertruyden, Pierre Van Damme, Hypolite Muhindo Mavoko. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 09.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Original Paper

Lateralization and Bodily Patterns of Segmental Signs and Spontaneous Pain in Acute Visceral Disease: Observational Study

Nour Shaballout[1*], MD, PhD; Anas Aloumar[1,2*], MD; Jorge Manuel[1,3], MSc; Marcus May[1], MD; Florian Beissner[1,4], MSc, PhD

[1]Somatosensory and Autonomic Therapy Research, Institute for Neuroradiology, Hannover Medical School, Hannover, Germany

[2]Department of Internal Medicine, Klinikum Region Hannover, Großburgwedel, Hannover, Germany

[3]Institute of Aerospace Medicine, German Aerospace Centre, Cologne, Germany

[4]Insula Institute for Integrative Therapy Research, Hannover, Germany

[*]these authors contributed equally

Corresponding Author:
Florian Beissner, MSc, PhD
Insula Institute for Integrative Therapy Research
Brabeckstraße 177e
Hannover, 30539
Germany
Phone: 49 16095543423
Email: f.beissner@insula-institut.org

Abstract

Background: The differential diagnosis of acute visceral diseases is a challenging clinical problem. Older literature suggests that patients with acute visceral problems show segmental signs such as hyperalgesia, skin resistance, or muscular defense as manifestations of referred visceral pain in somatic or visceral tissues with overlapping segmental innervation. According to these sources, the lateralization and segmental distribution of such signs may be used for differential diagnosis. Segmental signs and symptoms may be accompanied by spontaneous (visceral) pain, which, however, shows a nonsegmental distribution.

Objective: This study aimed to investigate the lateralization (ie, localization on one side of the body, in preference to the other) and segmental distribution (ie, surface ratio of the affected segments) of spontaneous pain and (referred) segmental signs in acute visceral diseases using digital pain drawing technology.

Methods: We recruited 208 emergency room patients that were presenting for acute medical problems considered by triage as related to internal organ disease. All patients underwent a structured 10-minute bodily examination to test for various segmental signs and spontaneous visceral pain. They were further asked their segmental symptoms such as nausea, meteorism, and urinary retention. We collected spontaneous pain and segmental signs as digital drawings and segmental symptoms as binary values on a tablet PC. After the final diagnosis, patients were divided into groups according to the organ affected. Using statistical image analysis, we calculated mean distributions of pain and segmental signs for the heart, lungs, stomach, liver/gallbladder, and kidneys/ureters, analyzing the segmental distribution of these signs and the lateralization.

Results: Of the 208 recruited patients, 110 (52.9%) were later diagnosed with a single-organ problem. These recruited patients had a mean age of 57.3 (SD 17.2) years, and 40.9% (85/208) were female. Of these 110 patients, 85 (77.3%) reported spontaneous visceral pain. Of the 110, 81 (73.6%) had at least 1 segmental sign, and the most frequent signs were hyperalgesia (46/81, 57%), and muscle resistance (39/81, 48%). While pain was distributed along the body midline, segmental signs for the heart, stomach, and liver/gallbladder appeared mostly ipsilateral to the affected organ. An unexpectedly high number of patients (37/110, 33.6%) further showed ipsilateral mydriasis.

Conclusions: This study underlines the usefulness of including digitally recorded segmental signs in bodily examinations of patients with acute medical problems.

(J Med Internet Res 2021;23(8):e27247) doi:10.2196/27247

KEYWORDS

digital pain drawings; visceral referred pain; referred pain; head zones; mydriasis; chest pain; clinical examination; differential diagnosis; digital health; digital drawings; pain; health technology; image analysis

Introduction

The differential diagnosis of acute visceral diseases is a common but challenging clinical problem. Since pain originating from visceral organs (ie, visceral pain) often exhibits characteristic patterns [1-8], many textbooks assign pain location a discriminative role in the differential diagnosis [9,10]. However, many studies have also reported negative results when testing the predictive power of pain location [11,12]. For example, pain localization in patients with coronary heart disease does not significantly differ from chest pain patients without coronary heart disease [13].

While primary visceral pain itself is a poorly defined, midline sensation, it starts to be referred or “transferred” to somatic structures when it persists for several minutes or longer [14,15]. These somatic structures can include skin, subcutaneous tissue, and muscle and are characterized by an overlapping segmental innervation with that of the diseased organ [16-29]. In these instances, referred visceral pain manifests as hyperalgesia, a phenomenon first described by Ross and Sturge in the 1880s [30,31] and subsequently studied in depth by Head and Mackenzie [16-19]. Head mapped out the cutaneous zones of referred hyperalgesia for all major organs and compared them with the location of skin lesions in herpes zoster [16]. The result is still considered one of the most precise maps of segmental innervation [32,33].

To the present day, zones of referred hyperalgesia in visceral disease carry Head’s name in many European countries, such as France, Germany, and Spain. In other parts of the world, however, clinicians mainly speak of “dermatomes,” and clinicians hardly know the term “Head zones” as well as Head’s work, in general. Some authors have even called segmental anatomy a “wrongly forgotten science” [29]. Only rarely do clinicians know that the transmitted signs are not limited to hyperalgesia of the skin but instead show a plethora of manifestations, including sensory disturbances such as allodynia and deep hyperalgesia (ie, Mackenzie zones); motor disturbances such as increased resistance of the skin, muscular defense, and resistance to passive joint movement; and, finally, signs of sympathetic activation such as vasomotor changes, asymmetric hyperhidrosis (ie, asymmetric sweating between left and right side of body), piloerection, and anisocoria (ie, unequal pupil size). As such, they are not limited to the dermatomes but include the myotomes, sclerotomes, and other parts of the segmental innervation [4]. Furthermore, segmental signs may be accompanied by symptoms of viscero-visceral reflexes such as nausea, vomiting, diarrhea, constipation, meteorism, and urinary retention [33].

To our knowledge, a systematic evaluation of simultaneously collected segmental signs and symptoms in patients has never been published in the English scientific literature. In Germany, however, Karl Hansen (1893-1962) and Hans Schliack (1919-2008) had studied a wide variety of segmental signs over several decades. While their results have only been published in German [33], the essence of their work has recently been made available in book form and extended by the work of other clinicians [29]. In a large sample of internal medicine patients, Hansen and Schliack [33] confirmed many of Head’s observations and greatly extended them to include all of the above-mentioned segmental signs and symptoms. Even more than Head, the authors emphasized the importance of sign lateralization by defining a side rule, according to which segmental signs are most likely to appear ipsilateral to the affected organ (Table 1).

Table 1. Lateralization of segmental signs for individual organs according to Hansen and Schliack [33].

Organ	Segmental signs by side of body and part of organ (yes, possible,[a] or no[b])	
	Right	Left
Heart	Possible	Yes
Pericardium	No	Yes
Aorta	Possible	Yes
Lung and bronchi	Yes	Yes
Pleura	Yes	Yes
Stomach	Yes (pylorus)	Yes (corpus, fundus)
Small intestine	Yes (duodenum, ileum)	Yes (jejunum)
Pancreas	No	Yes
Liver	Yes	No
Gallbladder	Yes	No
Spleen	No	Yes
Large intestine	Yes (caecum, appendix, ascending colon, proximal part of transverse colon)	Yes (distal part of transverse colon, descending colon, sigmoid colon, rectum)
Kidney	Yes	Yes
Ureter	Yes	Yes
Testis and ovary	Yes (testis, ovary, salpinx)	Yes (testis, ovary, salpinx)

[a] Possible indicates a possible but unlikely occurrence of signs from that organ.

[b] No indicates that segmental signs from a particular organ were never observed on that side.

A methodological problem that has hampered clinical research of segmental signs for a long time is the difficulty in adequately measuring bodily signs. However, recent developments in the field of digital pain drawings offer new and exciting possibilities to systematically record not only pain sensations but also segmental signs and analyze them using methods of statistical image analysis [34].

Here, we report the results of a study that investigated both the bodily patterns and lateralization of segmental signs and spontaneous pain in acute visceral diseases. We aimed to derive mean distributions of spontaneous pain and segmental signs for as many internal organs as possible and to analyze their segmental content and lateralization. To achieve this, we combined digital pain drawing technology and a structured, 10-minute bodily examination in a sample of emergency room patients.

Methods

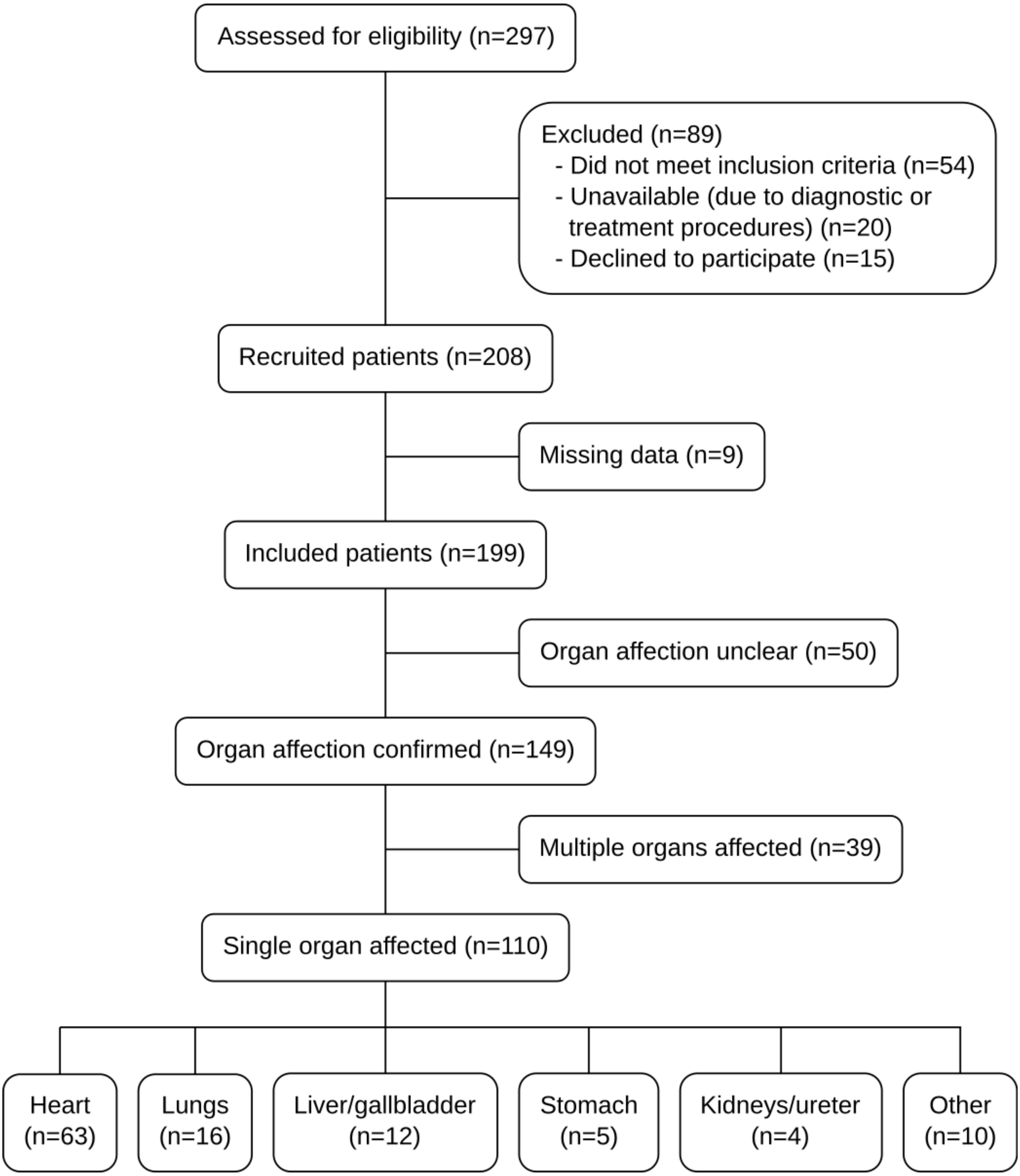
Ethics

The study was approved by the Ethics Committee of Hannover Medical School (number 2987-2017) and was conducted under the Declaration of Helsinki. All patients provided written informed consent after they were informed about the purpose of the study.

Study Population

Our study population consisted of patients from the emergency department of Hannover Medical School who were referred to internal medicine physicians between March 2017 and October 2017. Eligible patients were adults (age ≥18 years in Germany), presenting with an acute medical problem and with the ability to provide written informed consent. Furthermore, patients needed to be oriented as to place, time, and person. Exclusion criteria comprised refusal or inability to provide written consent, previously known or acutely diagnosed spinal cord injury, pregnancy, acute or past ocular illnesses, acute or past central or peripheral nervous disease, uncooperative patients, and patients who only presented to the emergency room for educational purposes or to receive a prescription. For a flowchart, see Figure 1.

Figure 1. Flowchart of the study.


Assessed for eligibility (n=297)
Excluded (n=89)
- Did not meet inclusion criteria (n=54)
- Unavailable (due to diagnostic or treatment procedures) (n=20)
- Declined to participate (n=15)
Recruited patients (n=208)
Missing data (n=9)
Included patients (n=199)
Organ affection unclear (n=50)
Organ affection confirmed (n=149)
Multiple organs affected (n=39)
Single organ affected (n=110)
Heart (n=63)
Lungs (n=16)
Liver/gallbladder (n=12)
Stomach (n=5)
Kidneys/ureter (n=4)
Other (n=10)

We recruited 208 patients (85, 40.9% women) for participation in our study. Nine drawings were lost due to technical failure of a tablet PC during the physical examination. The characteristics of the final study population can be found in Table 2, and their final diagnoses can be found in Multimedia Appendix 1.

Table 2. Demographics of the study population (n=208).

Characteristic	Value
Age (years), mean (SD)	57.3 (17.2)
Age range, n (%)	
18-39	34 (17.1)
40-59	66 (33.2)
60-79	80 (40.2)
≥80	19 (9.5)
Gender, n (%)	
Women	85 (40.9)
Men	123 (59.1)
Main complaint, n (%)	
Chest pain	88 (44.2)
Abdominal pain	55 (27.6)
Dyspnea	22 (11.1)
Other	34 (17.1)

Procedures

All clinical data were collected by 2 of the authors (AA and NS), henceforth called examiners. AA is an internal medicine specialist, and NS is a physician with 4 years of training for an internal medicine specialization. The examiners were fully informed about the study purpose and trained to do the physical examination for segmental signs and symptoms according to the protocol described below. Prior to the study, the examiners trained intensively together to ensure their physical examinations were standardized. This was also necessary to ensure that all procedures could be completed in a very limited timeframe.

During recruitment, the examiners screened the emergency dashboard to identify patients who were referred to internal medicine specialists. They approached all eligible patients, informed them about the study, and obtained written informed consent.

The examination took place directly after triage and before any medical intervention, diagnosis, or treatment. The examination lasted between 7 and 15 minutes, depending on the patient’s compliance (ie, general motivation to be examined, speed of undressing and answering the examiner’s questions, precision of the answers given, unrelated conversation, etc) and interruptions by nurses and physicians (as routine diagnostics and medical interventions had priority over the scientific investigation). Directly after the physical examination, all findings were recorded on a tablet computer running the app “SymptomMapper” (described in the section “Tablet Computer and Software Application”).

Categories of Findings

The clinical findings we were interested in can be divided into 3 groups, according to the ways they were recorded in the tablet: (1) distributed findings (ie, those with a bodily pattern), (2) lateralized findings (ie, those without a bodily pattern but with clear lateralization), and (3) other findings.

First, distributed findings were spontaneous pain, allodynia, superficial hyperalgesia, deep hyperalgesia, superficial skin resistance, muscle resistance, defense, asymmetric hyperhidrosis, piloerection, vasomotor changes, herpes zoster, and resistance to passive movement of the limbs. Distributed findings were recorded by the examiners in the form of electronic drawings on a body template, thus capturing their exact location and extent.

Second, lateralized findings were anisocoria, glossy eye, eyelid separation, tense facial muscles, asymmetric posture, and reduced respiration movements. These findings were recorded by choosing from a list of the abovementioned findings in conjunction with a side label (eg, “glossy eye right,” “mydriasis left,” etc).

Third, other findings were symptoms potentially related to viscero-visceral reflexes, namely, nausea, vomiting, constipation, diarrhea, meteorism, and urinary retention. These findings were recorded by choosing from a simple list of the abovementioned symptoms.

Tablet Computer and Software Application

All findings were recorded on Galaxy Note 10.1 (2014) tablet PCs with an electronic stylus based on inductive digitizing technology (Samsung, Seoul, South Korea). The tablets had a 10.1-inch touch screen with a resolution of 800×1280 pixels and were running Android 5.1.1 (Open Handset Alliance, Mountain View, CA, United States). The stylus was used for all data entry, hence allowing for a higher resolution while eliminating unwanted activation of the screen, for example, by the palm. The tablet and stylus were disinfected after every patient using disinfectant wipes.

We used a modified version of the SymptomMapper app developed by our group (Somatosensory and Autonomic Therapy Research, Institute for Neuroradiology, Hannover Medical School, Hannover, Germany) to acquire electronic pain drawings [35]. Its usability for doctors and the reliability of its symptom-drawing approach have recently been shown [35]. The app allowed the examiners to enter all findings from the bodily examination quickly. They could either draw distributed findings on a body template or choose from a list of lateralized or other findings. For the electronic drawings, examiners had a front and back view of the body available, and each newly added sign or symptom was displayed in a semitransparent way and in a different color.

Bodily Examination

Distributed Segmental Sign Examination

Our approach to the bodily examination was based entirely on Hansen and Schliack (p140-176) [33]. Its primary purpose was to check for the presence and record the extent and lateralization of pain and segmental signs and symptoms. We start by describing how distributed segmental signs were collected, as this was the same for different body regions (see below). These collection methods were (1) visual inspection and (2) palpation.

First, for visual inspection, the skin was visually inspected for the following signs: shingles (as a potential sign of Zoster reactivation), vasomotor changes (ie, skin color changing to red, pale, or blue, as a sign of sympathetic reflexes); piloerection (ie, any hair erection or “goosebumps,” as a sign of sympathetic reflexes), and muscular asymmetries (eg, asymmetric posture, tense facial muscles, respiratory chest movement, etc).

Second, for palpation, the body was palpated with warm hands to test for the following signs of sympathetic reflexes, increased muscle tone, or sensory disturbance: (1) asymmetric hyperhidrosis, (2) superficial hyperalgesia, (3) deep hyperalgesia, (4) allodynia, (5) superficial skin resistance, and (6) muscle resistance. Among palpation, first, for asymmetric hyperhidrosis, the skin was observed and palpated for any local differences in the amount of sweating. Second, for superficial hyperalgesia, the patient was informed that the examination could cause a little twinge and then was asked if the tip of a neurological examination needle (Healthstar, Lakewood, NJ, United States), when passed vertically over the skin in long and slow strokes, caused a different sensation in any area. Third, for deep hyperalgesia, folds of skin were held gently between the thumb and index finger or the region was tapped on. The test was considered positive if this procedure caused dull pain that lasted longer in some part of the body than in other parts of the body. Fourth, for allodynia, patients were asked if their clothes caused an unpleasant sensation somewhere on the body. Then, they were asked if a medical cotton swab passed over the skin in long and slow strokes caused a different sensation in any area. Fifth, superficial skin resistance was tested by superficial palpation of the trunk skin using the palm with very soft pressure. If the examiner felt either resistance or a rubbery membrane in any area, the test was considered positive in this area. Sixth, for muscle resistance, deep palpation of the trunk wall was performed on the front and back sides with the palm to detect the guarding of the trunk’s wall muscles (ie, anterior thoracic muscles, anterolateral abdominal wall muscles, posterior superficial muscles, and posterior deep muscles).

Complete Examination

The complete examination program had the following steps carried out in the exact order specified here: examination of (1) asymmetric posture, (2) pain and segmental symptoms, (3) the head, (4) the neck and chest, (5) the abdomen, and (6) the limbs.

First, for asymmetric posture, a general visual inspection of the patient’s posture was carried out directly after entering the examination room to check side differences in muscle tone.

Second, pain and segmental symptoms were collected by asking patients the following questions: (1) “Do you have pain—where, exactly? Do you have a headache?” and, when the patient reported pain, the painful region was drawn; (2) “Do you have nausea? Did you vomit since the onset of symptoms?”; (3) “Do you have diarrhea or constipation?”; (4) “Do you feel that your abdomen is full of gas?”; and (4) “Did you have any problem with urination since the onset of symptoms?”

Third, the head was examined with (1) special tests and (2) tests of distributed segmental signs. For the head, first, the special tests examined (1) the pupils, (2) the eyes/eyelids, and (3) tense facial muscles. For these special tests, first, for pupils, we tested for mydriasis, a sign of sympathetic activation, by equally exposing both eyes to light after instructing the patient to relax and look far away. The examiner used one hand to shadow the eyes and compared pupil diameters on both sides. This was repeated 3-5 times. In the case of a striking side difference, the test was considered positive for mydriasis. Second, for eyes/eyelids, eyelid separation (due to eyelid retraction) and eye gloss (due to excessive lacrimation) are both signs of sympathetic activation and were assessed by visually comparing the visible area and gloss of both eyes. In the case of a striking side difference, the more open and glossier eye was noted. Third, for tense facial muscles, a potential asymmetry of facial features caused by side differences in muscle tone was checked visually. It was considered positive when the upper lip was noticeably higher, the nasolabial fold deeper, and the cheeks more retracted on one side than on the other. The test was repeated once under provocation by applying pressure with the index and middle fingers on a point between the 2 heads of the sternocleidomastoid muscle. In terms of head tests overall, second, the distributed segmental signs were tested, including zoster, vasomotor changes, piloerection, asymmetric hyperhidrosis, superficial hyperalgesia, allodynia, and superficial skin resistance.

Fourth, as part the complete examination, was the neck and chest, where the patient’s front was examined while the patient was in a supine position after freeing the chest from clothes. Then, the back was examined with the patient sitting or lying on one side. Similar as for the head, the neck and chest included special tests (ie, the patient's chest movement during inspiration and expiration was observed during the visual inspection over several respiratory cycles, and any striking side differences were noted as a sign of increased muscle tone) and distributed segmental sign tests (ie, tests for zoster, vasomotor changes, piloerection, asymmetric hyperhidrosis, superficial hyperalgesia, deep hyperalgesia, allodynia, superficial skin resistance, and muscle resistance).

Fifth, for the abdomen, the patient was examined on the front in a supine position after freeing the abdomen. Then, the back was examined with the patient sitting or lying on one side. Again, special tests were performed; mainly, the defense was examined by applying sudden deep palpation over the painful areas of the abdomen. If the examiner felt a reflex of the abdominal wall, it was considered positive. Next, distributed segmental signs were tested: zoster, vasomotor changes, piloerection, asymmetric hyperhidrosis, superficial hyperalgesia, deep hyperalgesia, allodynia, superficial skin resistance, and muscle resistance.

Sixth, for limbs, as for the other body parts, there were special tests and testing of distributed segmental signs. For special tests, passive movements of the joints were examined to detect any resistance due to increased muscle tone. The distributed segmental signs test included zoster, vasomotor changes, piloerection, superficial hyperalgesia, and allodynia.

Patient Selection

Medical reports of all recruited patients were followed up through Hannover Medical Schools' electronic health records by 3 of the authors (NS, AA, and MM), to identify those patients with a definite diagnosis of visceral disease. All information regarding the acute complaint, previous diagnoses, and diagnostic procedures (electrocardiogram, laboratory, radiology, etc) were reviewed, and the most likely etiology for each patient was discussed. Patients without a definite visceral diagnosis were excluded from further analysis (n=50). The remaining cases were divided into those where a single organ was affected (n=110) and those with multi-organ problems (n=39). Only the single-organ cases were included in the final analysis, and only organs with at least 4 patients in the sample were included in any organ-specific analyses (Figure 1).

Data Analysis

General Considerations on Lateralization

According to Hansen and Schliack [33], the majority of segmental signs are lateralized and appear on specific sides of the body defined by the innervation of the individual organs (Table 1). In particular, the lateralization of signs for paired organs such as lungs and kidneys depends on which side is affected. Due to the nature of this study, it was not possible to conduct separate analyses for the left and right side in diseases of the lungs and kidneys/ureters. Furthermore, many lung cases were bilateral affections. Information about the lateralization of segmental signs for lungs and kidneys/ureters is, therefore, of little value and only shown for the sake of completeness.

Lateralization and Other Findings

We extracted all lateralized findings (ie mydriasis, glossy eye, eyelid separation, tense facial muscles, asymmetric posture, and reduced respiration movements) and other findings (nausea, vomiting, constipation, diarrhea, meteorism, and urinary retention) from SymptomMapper's JavaScript Object Notation files using a custom-written Python script (Python 2.7, Python Software Foundation, 2018). Then, we calculated, for each organ and each finding, the percentage of patients that had that finding. For lateralized findings, we calculated the percentage for each side individually, treating front and back as one surface. Finally, we calculated the mean frequency of each finding (ie, how often it was observed), irrespective of the specific organ.

Distributed Findings

Digital drawings from the app were converted to Nifti format (Neuroimaging Informatics Technology Initiative, 2017) with a custom-written Python script (Python 2.7, Python Software Foundation, 2018) and analyzed using tools from the Functional Magnetic Resonance Imaging (FMRIB) Software Library (FSL) version 5.0 (FMRIB Analysis Group, Oxford University, United Kingdom). Figures were prepared using VINCI (Volume Imaging in Neurological Research, Co-Registration and Region of Interest (ROI)s Included) 4.86.0 (Max Planck Institute for Metabolism Research, Cologne, Germany) and GNU Image Manipulation Program (GIMP; version 2.8.16, The GIMP Team).

First, to derive the bodily distribution of all segmental signs, all distributed signs were superimposed and the result binarized. In the resulting map, a pixel of value 1 on the body template meant that at least 1 sign had been found at that particular point on the body, in that particular patient. Binarization meant that we disregarded the number of signs that each patient showed and instead only considered their bodily location.

We then analyzed distributed signs individually to assess the segmental distribution for each sign according to the segmental scheme of Hansen and Schliack [33], which is largely based on Head's scheme [14,32]. To do this assessment, we calculated, for each segment, the percentage of the segment covered by the sign. For this calculation, we divided the pixel count by the total number of pixels of the respective segment. Only segments with at least 5% coverage were included. This arbitrary threshold was set to ensure that segments with marginal coverage (eg, due to drawing imperfections) were excluded. To assess the lateralization of findings, we further divided segments into left and right body halves, calculating the percentage for each of them. This resulted in a list of half segments covered by each sign. Finally, we calculated, for each organ, the mean number of segmental signs per half segment and the mean frequency of each sign across all organs.

Spontaneous pain was analyzed in the same way but separately from all other signs.

Results

The Overall Frequency of Signs and Symptoms

Of the 110 patients in our final sample, 85 (77.3%) had spontaneous pain, 81 (73.6%) showed at least 1 segmental sign, and 52 (47.3%) showed at least 1 segmental symptom. On average, each patient had a mean of 1.80 (SD 1.86) segmental signs and 0.77 (SD 1.00) segmental symptoms. The most frequent signs and symptoms are shown in Table 3.

Table 3. Frequency of segmental signs and symptoms in our patient sample (N=110).

Segmental sign or symptom	Value, n (%)
Segmental sign	
Superficial hyperalgesia (Head zone)	46 (41.8)
Muscle resistance	39 (35.5)
Mydriasis	37 (33.6)
Defense	13 (11.8)
Deep hyperalgesia (Mackenzie zone)	13 (11.8)
Superficial skin resistance	12 (10.9)
Tense facial muscles	11 (10.0)
Vasomotor changes	10 (9.1)
Glossy eye/wide eyelid	8 (7.3)
Asymmetric posture	7 (6.4)
Reduced respiration movements	3 (2.7)
Allodynia	2 (1.8)
Piloerection	1 (0.9)
Asymmetric hyperhidrosis	1 (0.9)
Zoster	0 (0)
At least 1 segmental sign	84 (76.4)
Segmental symptoms	
Nausea	45 (40.9)
Vomiting	18 (16.4)
Diarrhea	10 (9.1)
Meteorism	8 (7.3)
Constipation	5 (4.5)
Urinary retention	0 (0)
At least 1 segmental symptom	52 (47.3)
Spontaneous pain	85 (77.3)

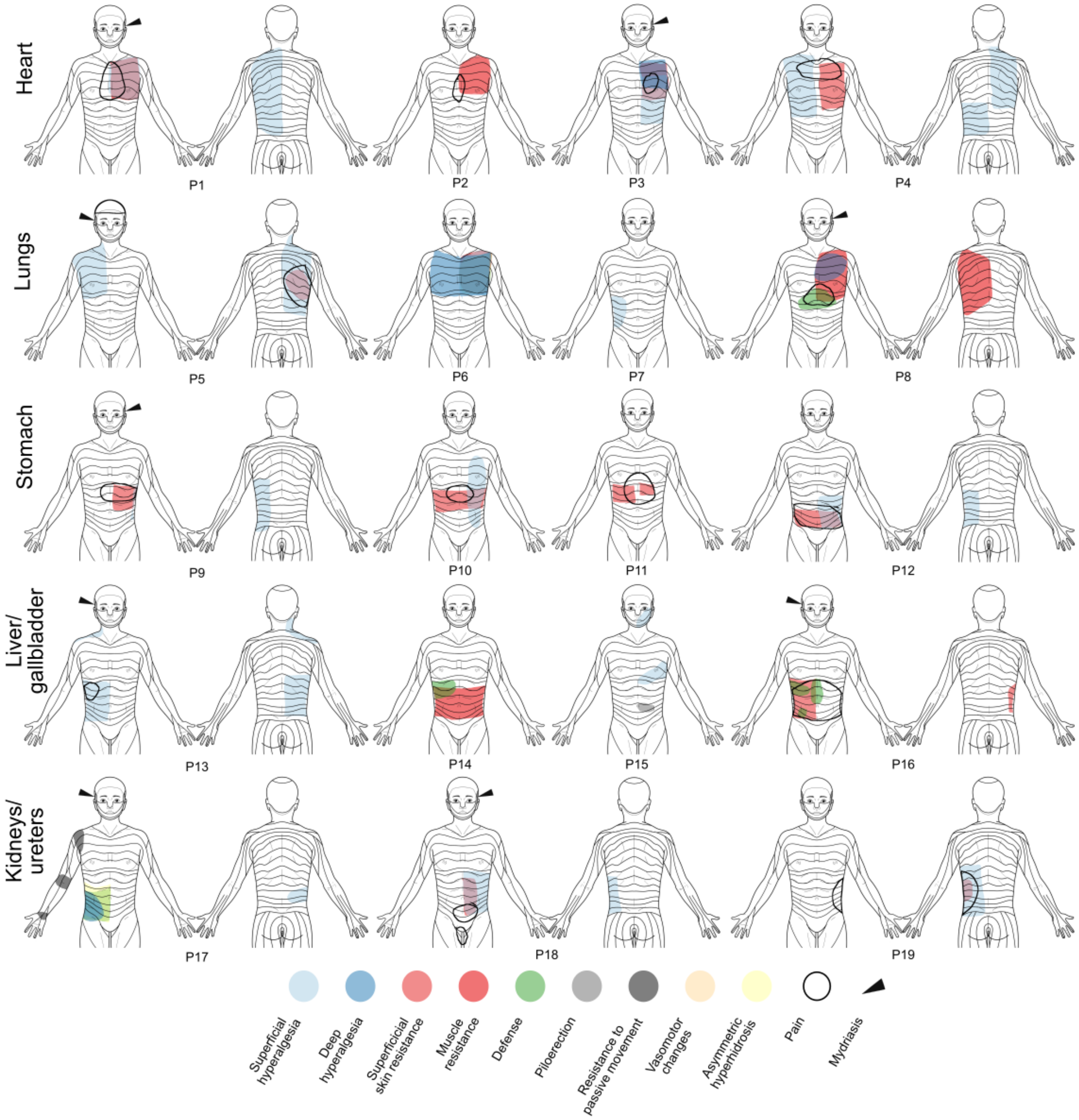
Frequency of Lateralization of Signs and Symptoms

All lateralization of signs and segmental symptoms are shown in Multimedia Appendix 2. As predicted by the side rule (see [33] and Table 1), the majority of lateralization signs were ipsilateral to the affected organ for the unpaired organs heart, stomach, and liver/gallbladder. The most striking finding was the high number of patients showing ipsilateral mydriasis as a potential sign of unilateral sympathetic activation. This lateralization was 100% ipsilateral for diseases of the liver/gallbladder (5 right vs 0 left), 100% ipsilateral for stomach diseases (1 left vs 0 right), and 83% ipsilateral for heart diseases (15 left vs 3 right).

Segmental Signs and Spontaneous Pain in Individual Patients

Bodily maps of segmental signs and spontaneous pain for a representative selection of individual patients are shown in Figure 2. The cases shown in Figure 2 reflect the entire bandwidth of segmental signs encountered in patients presenting to the emergency room. It ranges from “textbook cases” (eg, patients 2, 3, 6, 9, 13, 18, and 19), where segmental signs alone allow for a preliminary diagnosis, to those where segmental signs are hardly helpful or even misleading (eg, patients 7, 12, and 15). Their primary diagnoses and demographic information are summarized in Multimedia Appendix 3.

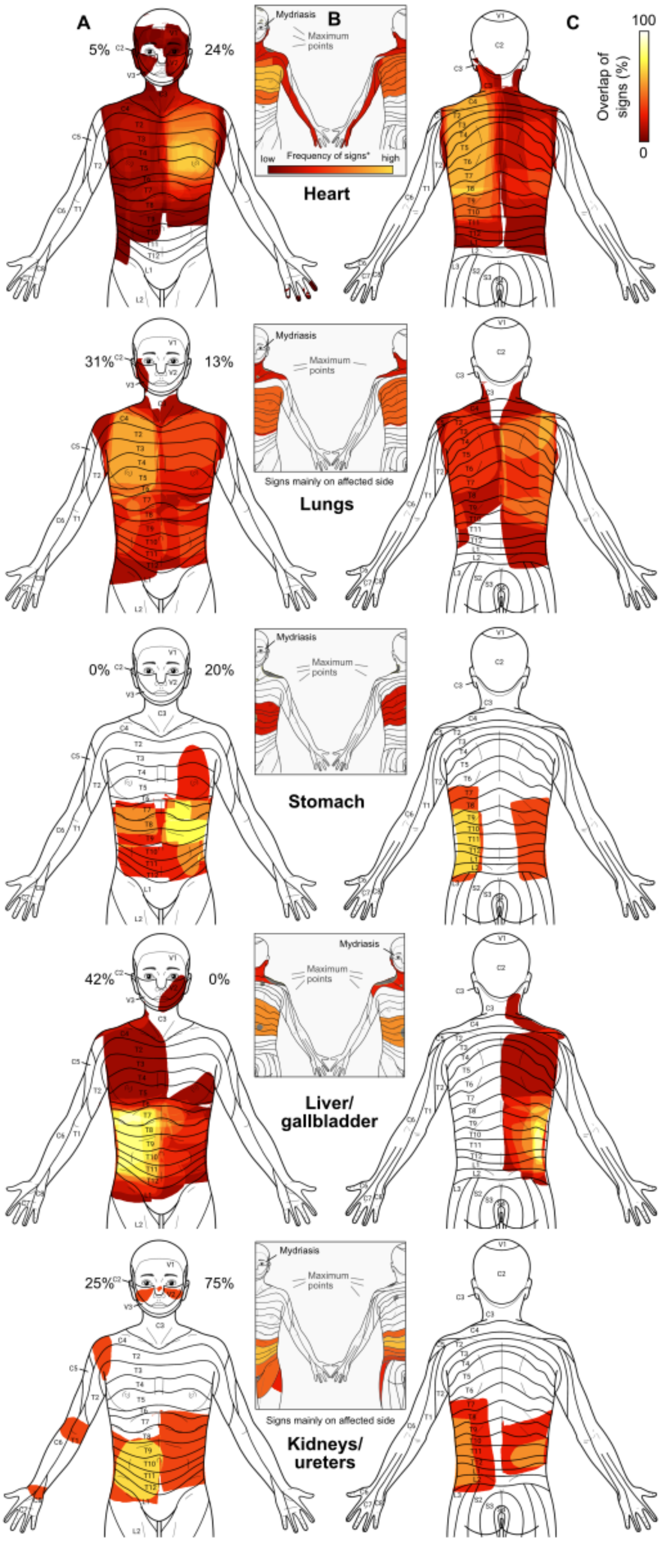
Figure 2. Segmental signs and spontaneous pain in individual patients with acute visceral diseases. P: patient.


Heart
Lungs
Stomach
Liver/gallbladder
Kidneys/ureters
P1
P2
P3
P4
P5
P6
P7
P8
P9
P10
P11
P12
P13
P14
P15
P16
P17
P18
P19
Superficial hyperalgesia
Deep hyperalgesia
Superficial skin resistance
Muscle resistance
Defense
Piloerection
Resistance to passive movement
Vasomotor changes
Asymmetric hyperhidrosis
Pain
Mydriasis

Bodily Maps and Segmental Patterns of Distributed Signs

Bodily maps of all distributed segmental signs are shown in Figure 3, while Figure 4 contains detailed segmental information concerning the distribution of the individual signs and spontaneous pain. In general, the observed distributions of segmental signs were largely consistent with those reported by Hansen and Schliack [33]. The lungs were the only exception, which showed a more widespread distribution than predicted. Concerning lateralization, segmental signs from the unpaired organs showed a clear side difference, with more signs appearing ipsilateral to the affected organ, thus supporting the "side rule" represented in Table 1. For the lungs and kidneys/ureters, however, this rule could not be tested, since results for these organs reflected a mixture of left, right, and bilateral affections.

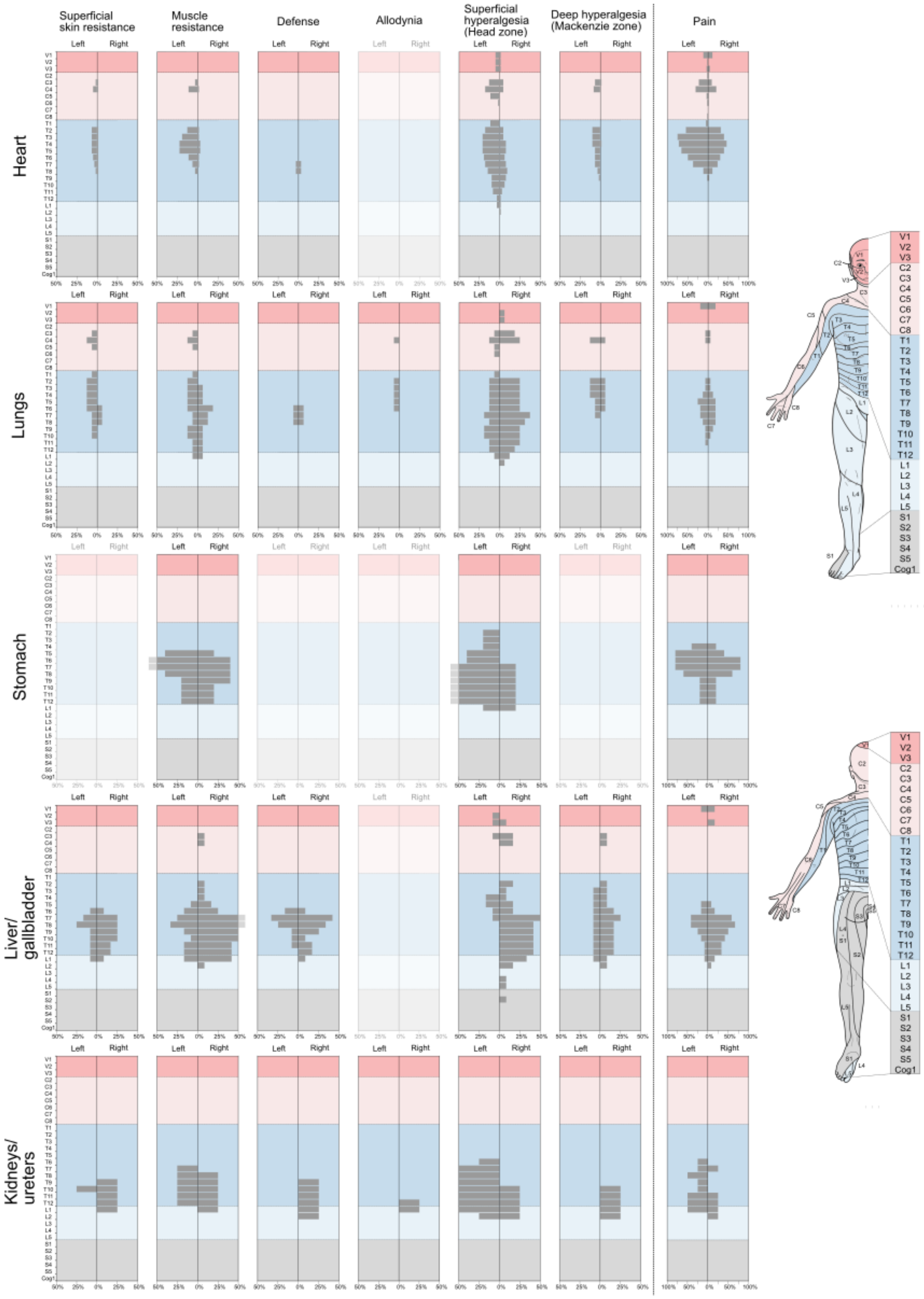
Figure 3. Distributed segmental signs in acute visceral diseases. Columns A and C show a front and back body map of all distributed segmental signs. The inserts in column B show the segmental distributions for each organ as reported by Hansen and Schliack [33], for comparison. Percentage values at the sides of the head indicate the frequency of unilateral mydriasis in affections of the respective organ.


A
B
C
Overlap of signs (%)
100
0
Mydriasis
Maximum points
low
Frequency of signs*
high
5%
24%
Heart
31%
13%
Signs mainly on affected side
Lungs
0%
20%
Stomach
42%
0%
Liver/ gallbladder
25%
75%
Signs mainly on affected side
Kidneys/ ureters

Within-organ comparison showed that the different segmental signs but also spontaneous pain had a similar segmental distribution (Figure 4). Between organs, these distributions showed considerable overlap. Superficial hyperalgesia (Head zone) exhibited the greatest spread in terms of segments.

Regarding lateralization, segmental signs for the unpaired organs mostly obeyed the side rule, according to which signs should appear on the body half where the organ is located. Pain differed markedly in that respect and, instead, showed a rather symmetric pattern.

Figure 4. Segmental patterns and lateralization of segmental signs and pain in acute visceral diseases. Each row contains the results for an individual organ, while columns represent the different most-common segmental signs and pain, respectively. Graphs without any occurrence were blanked. Please note the different axis scaling for the latter. For the sake of clarity, segmental sections have been color-coded according to the body template shown on the right. This body template shows the different segments: trigeminal (V1-3), cervical (C2-8), thoracal (T1-12), lumbar (L1-5), sacral (S1-5), and coccygeal (Cog1).



For cardiac-related conditions, segmental signs were mostly located in the thoracic segments and, to a lesser extent, in cervical segments. Superficial hyperalgesia (Head zone) was also detected in the trigeminal segments. The maximum of the averaged signs was in the T3-T5 region (Figure 3B), as predicted by Hansen and Schliack [33]. In terms of lateralization, all distributed signs were strongly left-dominated with deep hyperalgesia (Mackenzie zone), showing complete left-lateralization (Figure 4). In heart patients, defense and allodynia were rare and nonexistent, respectively.

For diseases of the lungs, segmental signs were very widespread and covered a range from V2 to L2. Signs were generally less focused than for the heart, and no clear maximum was discernible. In this, the distribution deviated from Hansen and Schliack's [33], who reported T9 as the lower margin of segmental signs in lung diseases. As was to be expected due to the mixture of left, right, and bilateral organ diseases, no lateralization could be seen.

For the stomach, the segmental distribution was almost strictly thoracic, from T2 to T12, with a maximum at T6-9 on the front and on the back. Similar to the heart, superficial hyperalgesia for the stomach was lateralized to the left, as predicted by the side rule. The comparison with Hansen and Schliack [33] showed that the maximum of signs in T6-9 fell in the expected range in the front view. On the back, however, there was only a partial overlap, with Hansen and Schliack [33] predicting higher thoracic segments than were found in our study.

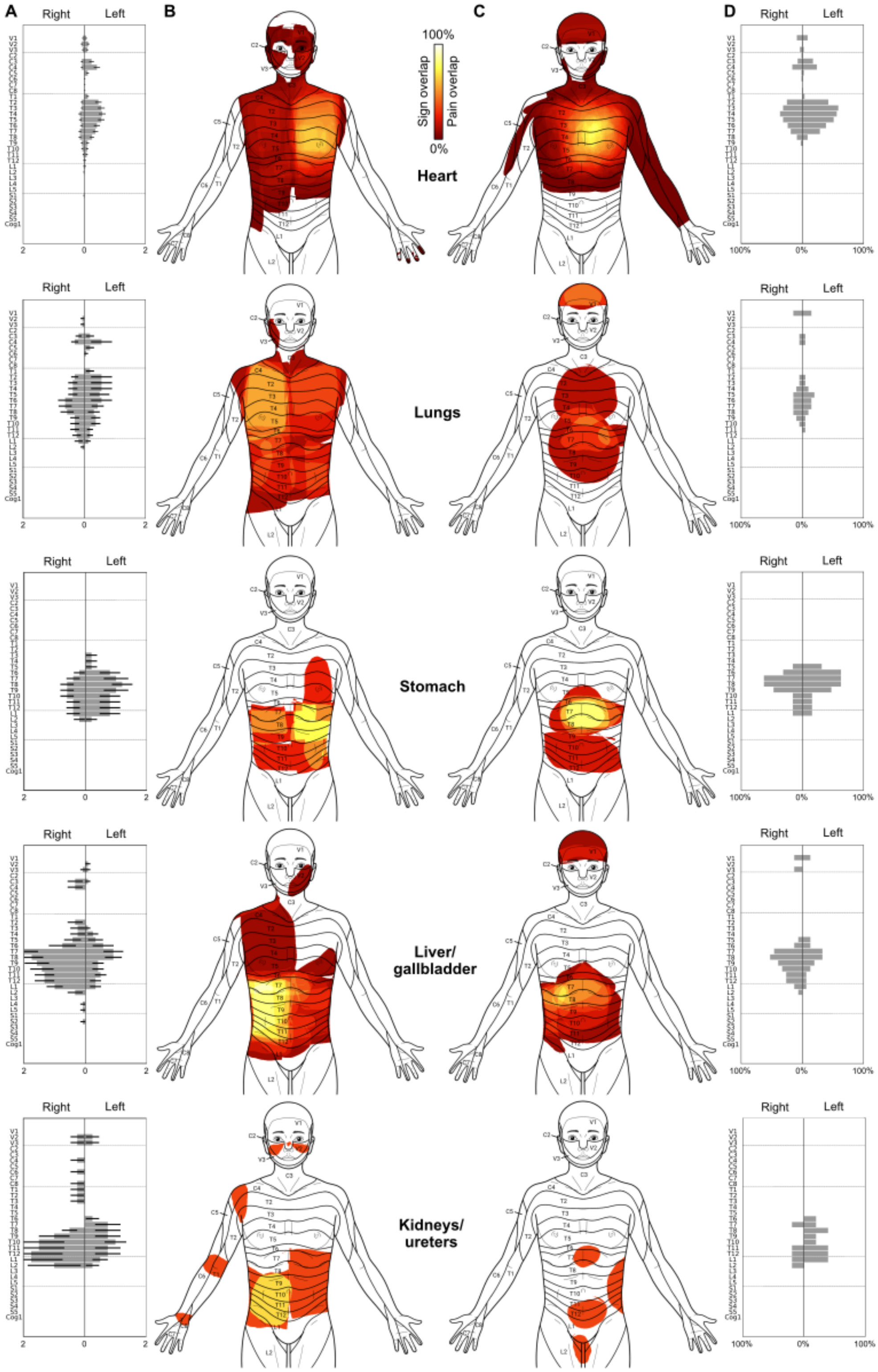
The similarity with Hansen and Schliack's [33] results was much higher for patients with liver/gallbladder diseases. Here, segmental signs showed a largely thoracic distribution but with the characteristic shoulder presentation in segments C3-C5 [33]. In terms of lateralization, muscle resistance, defense, and superficial and deep hyperalgesia were predominantly right-lateralized, with superficial hyperalgesia (Head zone) showing almost complete right-lateralization.

Finally, segmental signs of the kidneys had the narrowest distribution, starting at T6 and extending down to L2, once again, showing a rather high similarity with the predicted distribution by Hansen and Schliack [33].

Comparison of Spontaneous Pain and Segmental Signs

The segmental distributions of spontaneous pain and segmental signs are shown in Figure 5. It is evident that spontaneous pain differed markedly from segmental signs. It spanned fewer segments but extended to the head region (V1) in cardiac, respiratory, and liver/gallbladder affections. Furthermore, spontaneous pain was much less lateralized than segmental signs and, instead, rather localized in the body midline. In general, the pain was less widespread and showed much weaker lateralization than segmental signs in the unpaired organs (ie, heart, stomach, and liver/gallbladder).

Figure 5. Direct comparison between segmental signs (A and B) and spontaneous pain (C and D) in acute visceral diseases. Column A shows the mean number of segmental signs per segment for the individual organs, and column B shows the joint distribution of segmental signs (cf Figure 2). The symptom of spontaneous pain is shown as mean distributions of pain in column C, and their exact segmental content is shown in column D.



Discussion

Principal Findings

In this study, we investigated bodily patterns and lateralization of segmental signs and spontaneous pain in acute visceral diseases. We derived mean distributions of spontaneous pain and segmental signs for the heart, lungs, stomach, liver/gallbladder, and kidneys/ureters by combining digital pain drawing technology and a structured 10-minute bodily examination in patients presenting to the emergency room. We extracted precise information on the segmental content and lateralization and compared the results with the slightly outdated but authoritative German work of Hansen and Schliack [33]. Although purely descriptive by design, our study is the first in the English language to provide a detailed account of simultaneously collected segmental signs and symptoms for visceral diseases in the clinical setting.

Lateralization of Segmental Signs

The lateralization of segmental signs is important, as it allows one to quickly identify the affected body side (ie, the side hosting the affected organ; Table 1). This side rule may be useful in the differential diagnosis, such as in differentiating gastritis from hepatitis or pancreatitis, or acute coronary syndrome from a pulmonary embolism or esophagitis. Due to our study design and the very mixed patient sample, our results regarding lateralization were limited to the heart, stomach, and liver/gallbladder. Although segmental signs of the lungs and kidneys/ureters are also expected to be found ipsilateral to the affected side, a separate analysis for the individual sides was not possible for these organs due to the limited number of cases, many of which showed bilateral affections.

Although our data were not analyzed prospectively, it appears, for the heart, stomach, and liver/gallbladder, that they may support the findings of Hansen and Schliack [33] that segmental signs appear ipsilateral to the affected organ. While this was evident for the mean bodily distributions of segmental signs, we also found an ipsilateral occurrence of mydriasis, a finding rarely raised outside the neurological setting. It results from a reflex mediated by the ciliospinal center, which conducts impulses from the entire body to the sympathetically innervated dilator pupillae muscle (p271) [28] and, more than 100 years ago, was first described to occur in affections of the lungs [36] and the heart [37]. More recently, Rosenberg [38] has shown that anisocoria (ie, unequal pupil size) under physiological conditions is a manifestation of sympathetic asymmetry.

We found ipsilateral (ie, right-sided) mydriasis in 42% of our liver/gallbladder patients and not a single case of contralateral mydriasis. For the heart, mydriasis was less frequent (24% ipsilateral vs 5% contralateral), yet this means that patients showing the sign had it on the ipsilateral side in almost 83% of the cases. Hansen and Schliack [33] reported qualitatively similar but generally higher numbers for mydriasis. In their sample of 28 heart patients, 27 (96%) had mydriasis, and this was ipsilateral in 26 patients (96%). In 56 liver/gallbladder patients, 54 (96%) had mydriasis, of which 50 cases (93%) were ipsilateral (ie, right-sided).

The generally higher numbers of mydriasis in heart diseases found in Hansen and Schliack's [33] work may be explained by the fact that these authors used dark adaptation and infrared photographs in many of their patients, which made their examination less subjective, while our examiners were restricted to visual inspection under normal light. Clinicians interested in this phenomenon should consider using a portable infrared pupilometer.

Localization and Distribution of Segmental Signs

A subset of the findings collected in our study was further analyzed to extract detailed segmental information. We called this group of findings "distributed signs." It comprised a number of somatosensory (ie, superficial and deep hyperalgesia and allodynia), somatomotor (ie, superficial skin resistance, muscle resistance, and defense), and visceromotor signs (ie, vasomotor changes, piloerection, and asymmetric hyperhidrosis). Of these, superficial hyperalgesia (ie, Head zones), muscle resistance, defense, and deep hyperalgesia (ie, Mackenzie zones) were the most frequently observed in our sample of patients, while others, such as allodynia, piloerection, asymmetric hyperhidrosis, or zoster, were exceedingly rare.

There was a close similarity between the original maps of segmental signs by Hansen and Schliack [33] and our mean distributions of all signs (Figure 3). For a prospective evaluation, however, future studies should aim to quantify this similarity (eg, by using spatial similarity measures).

Several groups have studied individual segmental signs or groups of signs since the days of Hansen and Schliack. For example, Nicholas and colleagues [39] found that patients with myocardial infarction showed characteristic paravertebral soft tissue changes readily detected by palpation. Compared with patients without diagnosed cardiovascular diseases, patients with myocardial infarction had a significantly higher incidence of increased firmness, warmth, ropiness, oedematous changes, and heavy musculature, almost entirely confined to cardiac segments T1-4. In a follow-up 3 years after the infarction, these signs had regressed in the majority of patients [40]. Vecchiet and colleagues [41] found ipsilateral superficial and deep hyperalgesia of the first lumbar (L1) segment in patients after renal/ureteral calculosis.

For the gallbladder, Stawowy and colleagues [42] found that all patients with acute cholecystitis reported referred pain in the epigastrium and under the right curvature. Segmental signs inside this area were quantitatively evaluated using von Frey hairs, warm and cold metal rollers, and a constant current stimulator to test for the different forms of hypersensitivity or allodynia. The authors reported that 20% of the patients showed hypersensitivity or allodynia to mechanical, 53% to cold, 40% to warmth, and 63% to electrical stimulation [42]. The same authors reported that 50% to 56% of patients with acute appendicitis showed segmental signs over the right abdominal quadrant, with the maximum located approximately at McBurney point [43]. These findings were recently confirmed by Roumen and colleagues [44], who reported that 39% of patients with acute appendicitis demonstrated at least one segmental sign (ie, hyperalgesia, hypoesthesia, altered cool perception, or positive pinch test) over the lower right abdomen. Finally, a large number of smaller studies and case reports have been published, which have been reviewed by Beal [45].

Segmental Signs Versus Spontaneous Pain

The majority of our patients with visceral diseases reported spontaneous pain (Table 3). In 85% of the cases, it was, by far, the most frequent finding, followed by superficial hyperalgesia (46%), nausea (45%), and muscle resistance (39%).

Many textbooks assign pain location a discriminative role in the differential diagnosis (eg, retrosternal chest pain that radiates to the left arm or lower jaw usually refers to acute coronary heart disease. However, such predictive power of pain location has been a matter of debate for decades [11-13]. Here, we found, by direct comparison of spontaneous pain and segmental signs, that the two were rather dissimilar in their bodily patterns and segmental distributions (Figure 5). Irrespective of the affected organ, spontaneous pain was less widespread than segmental signs (ie, it included fewer segments). Furthermore, spontaneous pain appeared mostly in the body midline, thus lacking the diagnostically relevant ipsilateral distribution seen in the majority of segmental signs. As Figure 5 shows, patients with lung, stomach, and liver/gallbladder diseases all showed spontaneous pain in the epigastric region (T5-9), thus rendering this symptom unsuitable for differential diagnosis.

The substantial differences found between pain and segmental signs regarding their location and lateralization underline the importance of making a clear distinction between visceral pain, (referred) hyperalgesia, and other segmental signs.

Despite the purely descriptive design of this study, our results (Figures 4 and 5) regarding the benefit of using spontaneous pain or segmental signs seem to favor the latter over the former. Future studies should test this in a prospective way (eg, by letting a blinded assessors predict the affected organ from the distribution of spontaneous pain or from that of segmental signs).

Limitations

Our study had several limitations that need to be discussed. Firstly, our patient sample was relatively small, as we could only analyze approximately half of the included patients. There were two reasons for this. On the one hand, approximately one-quarter of our patients had to be excluded from the analysis, since they left the hospital without a confirmed final diagnosis. This is due to the unique situation in the emergency room, where the vital role of the specialist is to rule out life-threatening conditions. A further one-quarter of the remaining patients had to be excluded from the analysis because they suffered from diseases affecting multiple organs. Secondly, while we took great care to include only patients with single-organ problems, it is likely that affections of other organs were present but overlooked in some of the patients. This means that some patients who seemed to only present with cardiac disease may have had another underlying disease affecting other organs. Thirdly, findings collected by means of palpation are naturally more subjective than, for example, laboratory results. While there are ways to measure segmental signs more quantitatively, we did not do so to keep the examination time to an absolute minimum, as required by the clinical setting. Finally, we did not differentiate explicitly between signs and symptoms that patients had only during their acute problem from patient symptoms that occurred usually. This may have introduced some bias.

Conclusions

This study underlines the usefulness of including segmental signs in the bodily examination of patients with acute medical problems. As we have shown, capturing the location of segmental signs on a digital body map may assist in the clinical decision-making process in some acute visceral conditions. Segmental information and lateralization from the 3 most-frequent signs (superficial hyperalgesia, muscle resistance, and mydriasis) can be quickly acquired and may help physicians narrow the differential diagnosis.

Acknowledgments

The authors would also like to thank Christoph Duesberg, former Head of the Central Emergency Room at Hannover Medical School, for his support in planning and carrying out the study. The authors would like to thank the Horst Görtz Foundation for providing financial support.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary table 1.
[PDF File (Adobe PDF File), 42 KB - jmir_v23i8e27247_app1.pdf]

Multimedia Appendix 2
Supplementary figure 1.
[File , 125 KB - jmir_v23i8e27247_app2.]

Multimedia Appendix 3
Supplementary table 2.
[File , 54 KB - jmir_v23i8e27247_app3.]

References

1. Bennett J, Atkinson M. The Differentiation Between Œsophageal and Cardiac Pain. The Lancet 1966 Nov;288(7473):1123-1127. [doi: 10.1016/s0140-6736(66)92209-4]
2. Henderson RD, Wigle ED, Sample K, Marryatt G. Atypical chest pain of cardiac and esophageal origin. Chest 1978 Jan;73(1):24-27. [doi: 10.1378/chest.73.1.24] [Medline: 620553]

3. Berger J, Buclin T, Haller E, Van Melle G, Yersin B. Right arm involvement and pain extension can help to differentiate coronary diseases from chest pain of other origin: a prospective emergency ward study of 278 consecutive patients admitted for chest pain. J Intern Med 1990 Mar;227(3):165-172. [doi: 10.1111/j.1365-2796.1990.tb00138.x] [Medline: 2313224]
4. Giamberardino MA, Vecchiet L. Pathophysiology of visceral pain. Current Review of Pain 1997 Mar;1(1):23-33. [doi: 10.1007/bf02938147]
5. Yamamoto W, Kono H, Maekawa M, Fukui T. The relationship between abdominal pain regions and specific diseases: an epidemiologic approach to clinical practice. J Epidemiol 1997 Mar;7(1):27-32 [FREE Full text] [doi: 10.2188/jea.7.27] [Medline: 9127570]
6. Kalloo A. Overview of differential diagnoses of abdominal pain. Gastrointest Endosc 2002 Dec;56(6 Suppl):S255-S257. [doi: 10.1067/mge.2002.129004] [Medline: 12447278]
7. Bates C, Plevris J. Clinical evaluation of abdominal pain in adults. Medicine 2013 Feb;41(2):81-86. [doi: 10.1016/j.mpmed.2012.11.005]
8. Wilson SJ, MacLeod DC. Chest pain in acute medicine. Medicine 2013 Feb;41(2):113-117. [doi: 10.1016/j.mpmed.2012.11.010]
9. Kasper D, Fauci A, Hauser S, Longo D, Jameson J, Loscalzo J. Harrison's principles of internal medicine, 19e (Vol. In: 1, No. 2). Mcgraw-Hill. New York: McGraw Hill Education Medical; 2015.
10. Mahadevan S, Garmel G. An introduction to clinical emergency medicine. New York: Cambridge University Press; 2012.
11. Eriksson B, Vuorisalo D, Sylvén C. Diagnostic potential of chest pain characteristics in coronary care. J Intern Med 1994 May;235(5):473-478. [doi: 10.1111/j.1365-2796.1994.tb01105.x] [Medline: 8182404]
12. Everts B, Karlson BW, Währborg P, Hedner T, Herlitz J. Localization of pain in suspected acute myocardial infarction in relation to final diagnosis, age and sex, and site and type of infarction. Heart & Lung 1996 Nov;25(6):430-437. [doi: 10.1016/s0147-9563(96)80043-4]
13. Bösner S, Bönisch K, Haasenritter J, Schlegel P, Hüllermeier E, Donner-Banzhoff N. Chest pain in primary care: is the localization of pain diagnostically helpful in the critical evaluation of patients?--A cross sectional study. BMC Fam Pract 2013 Oct 18;14(1):154 [FREE Full text] [doi: 10.1186/1471-2296-14-154] [Medline: 24138299]
14. Ness T, Gebhart G. Visceral pain: a review of experimental studies. Pain 1990;41(2):167-234 https://www.sciencedirect.com/science/article/pii/0304395990900215. [doi: 10.1016/0304-3959(90)90021-5]
15. Giamberardino M, Vecchiet L. Visceral pain, referred hyperalgesia and outcome: new concepts. European journal of anaesthesiology.Supplement 1995;10:61. [doi: 10.1016/b0-12-226870-9/01739-1]
16. Head H. On Disturbances of Sensation With Especial Reference to the Pain of Visceral Disease. Brain 1893;16(1-2):1-133. [doi: 10.1093/brain/16.1-2.1]
17. HEAD H. On Disturbances of Sensation With Especial Reference to the Pain of Visceral Disease: Part II. Brain 1894;17(3):339-480. [doi: 10.1093/brain/17.3.339]
18. HEAD H. On Disturbances of Sensation With Especial Reference to the Pain of Visceral Disease: Part III. Brain 1896;19(2-3):153-276. [doi: 10.1093/brain/19.2-3.153]
19. Mackenzie J. Some Points Bearing on the Association of Sensory Disorders and Visceral Disease. Brain 1893;16(3):321-354. [doi: 10.1093/brain/16.3.321]
20. Sherren J. On the Occurence and Significance of Cutaneous Hyperalgesia in Appendictitis. The Lancet 1903 Sep;162(4177):816-821. [doi: 10.1016/s0140-6736(01)51715-8]
21. Elsberg CA, Neuhof H. The Diagnostic Value of Cutaneous Hyperalgesia (Head's Zones) in Abdominal Disease. The American Journal of the Medical Sciences 1908;135(5):690-711. [doi: 10.1097/00000441-190811000-00007]
22. Langstroth L. An Attempt to Determine the Diagnostic Importance of Head's Zones of Hyperalgesia. Arch Intern Med 1915 Aug 01;XVI(2):149. [doi: 10.1001/archinte.1915.00080020003001]
23. Bullowa JG. Segmental Hyperalgesia and Segmental Increased Muscle Tone in Diseases of the Lungs and Heart. Bull N Y Acad Med 1929 Jun;5(6):538-551 [FREE Full text] [Medline: 19311681]
24. Bolton C. Observations on Referred Pain. Brain 1934;57(3):211-226. [doi: 10.1093/brain/57.3.211]
25. Boas EP, Levy H. Extracardiac determinants of the site and radiation of pain in angina pectoris with special reference to shoulder pain. American Heart Journal 1937 Nov;14(5):540-554. [doi: 10.1016/s0002-8703(37)90278-3]
26. Brown FR. The Problem of Abdominal Pain. Br Med J 1942 May 02;1(4243):543-546 [FREE Full text] [doi: 10.1136/bmj.1.4243.543] [Medline: 20784206]
27. Kennard M, Haugen F. The relation of subcutaneous focal sensitivity to referred pain of cardiac origin. Anesthesiology: The Journal of the American Society of Anesthesiologists 1955;16(3):297-311. [doi: 10.1097/00000542-195505000-00001]
28. Schliack H. Klinische und theoretische Bedeutung der Headschen Zonen. Acta Neurovegetativa 1966 Mar;28(1-4):437-449. [doi: 10.1007/bf01227404]
29. Wancura-Kampik I. Segmental anatomy: the key to mastering acupuncture, neural therapy, and manual therapy. MUNICH, Germany: Urban & Fischer/Elsevier; 2012.
30. Sturge WA. The Phenomena of Angina Pectoris, and their Bearing Upon the Theory of Counter-Irritation. Brain 1883;5(4):492-510. [doi: 10.1093/brain/5.4.492]
31. Ross J. On the Segmental Distribution of sensory Disorders. Brain 1888;10(4):333-361. [doi: 10.1093/brain/10.4.333]

32. Lee M, McPhee R, Stringer M. An evidence-based approach to human dermatomes. Clin Anat 2008 Jul;21(5):363-373. [doi: 10.1002/ca.20636] [Medline: 18470936]
33. Hansen K, Schliack H. Segmentale Innervation: ihre Bedeutung für Klinik und Praxis. Stuttgart, Germany: Thieme; 1962.
34. Shaballout N, Neubert T, Boudreau S, Beissner F. From Paper to Digital Applications of the Pain Drawing: Systematic Review of Methodological Milestones. JMIR Mhealth Uhealth 2019 Sep 05;7(9):e14569 [FREE Full text] [doi: 10.2196/14569] [Medline: 31489841]
35. Neubert T, Dusch M, Karst M, Beissner F. Designing a Tablet-Based Software App for Mapping Bodily Symptoms: Usability Evaluation and Reproducibility Analysis. JMIR Mhealth Uhealth 2018 May 30;6(5):e127 [FREE Full text] [doi: 10.2196/mhealth.8409] [Medline: 29848470]
36. Roque. Note sur l'inégalité des pupilles dans les affections unilatérales des poumons. Gaz. méd (1869)., Paris 48 1869:639.
37. Gibson GA. Some Hitherto Undescribed Symptoms in Agina Pectoris. Brain 1905;28(1):52-64. [doi: 10.1093/brain/28.1.52]
38. Rosenberg ML. Physiologic Anisocoria: A Manifestation of a Physiologic Sympathetic Asymmetry. Neuro-Ophthalmology 2009 Jul 08;32(3):147-149. [doi: 10.1080/01658100802115254]
39. Nicholas AS, DeBias DA, Ehrenfeuchter W, England KM, England RW, Greene CH, et al. A somatic component to myocardial infarction. Br Med J (Clin Res Ed) 1985 Jul 06;291(6487):13-17 [FREE Full text] [doi: 10.1136/bmj.291.6487.13] [Medline: 3926040]
40. Nicholas AS, DeBias DA, Greene CH. Somatic component to myocardial infarction: three year follow up. BMJ 1991 Jun 29;302(6792):1581-1581 [FREE Full text] [doi: 10.1136/bmj.302.6792.1581] [Medline: 1855044]
41. Vecchiet L, Giamberardino M, Dragani L, Albe-Fessard D. Pain from renal/ureteral calculosisvaluation of sensory thresholds in the lumbar area. Pain 1989;36(3):e. [doi: 10.1016/0304-3959(89)90087-0]
42. Stawowy M, Bluhme C, Arendt-Nielsen L, Drewes AM, Funch-Jensen P. Somatosensory changes in the referred pain area in patients with acute cholecystitis before and after treatment with laparoscopic or open cholecystectomy. Scand J Gastroenterol 2004 Oct 08;39(10):988-993. [doi: 10.1080/00365520410003425] [Medline: 15513339]
43. Stawowy M, Rössel P, Bluhme C, Funch-Jensen P, Arendt-Nielsen L, Drewes AM. Somatosensory changes in the referred pain area following acute inflammation of the appendix. Eur J Gastroenterol Hepatol 2002 Oct;14(10):1079-1084. [doi: 10.1097/00042737-200210000-00008] [Medline: 12362098]
44. Roumen RMH, Vening W, Wouda R, Scheltinga MM. Acute Appendicitis, Somatosensory Disturbances ("Head Zones"), and the Differential Diagnosis of Anterior Cutaneous Nerve Entrapment Syndrome (ACNES). J Gastrointest Surg 2017 Jun 14;21(6):1055-1061. [doi: 10.1007/s11605-017-3417-y] [Medline: 28411350]
45. Beal M. Viscerosomatic reflexes: a review. J Am Osteopath Assoc 1985 Dec;85(12):786-801. [Medline: 3841111]

Edited by R Kukafka; submitted 19.01.21; peer-reviewed by T Ots, M Galve Villa, Y Cai; comments to author 13.02.21; revised version received 02.05.21; accepted 14.06.21; published 27.08.21.

Please cite as:
Shaballout N, Aloumar A, Manuel J, May M, Beissner F
Lateralization and Bodily Patterns of Segmental Signs and Spontaneous Pain in Acute Visceral Disease: Observational Study
J Med Internet Res 2021;23(8):e27247
URL: https://www.jmir.org/2021/8/e27247
doi:10.2196/27247
PMID:34448718

©Nour Shaballout, Anas Aloumar, Jorge Manuel, Marcus May, Florian Beissner. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org